


```
1497 AGCCTTAGTGAATCGGCTTGTCAGCGGAGGAGCGGTGCAACTGA 1546
|||||
462 nAlaPheSerGlnValGlyIleValSerGlyArgGlyThrIleuValLeuA 479
1547 ATGCCGATATACAGTTCACCCCGACAACTCTATTCGGCTTGGCGC 1596
|||||
479 snSerSerAsnGlnIleAsnProAspAsnIleuTyPheGlyPheArgGly 495
1597 GGAGCTTGAGTGAACGGCATTCGCTTCGTTCCACCGCATTCGAAA 1646
|||||
496 GlyArgIleuAspAlaAsnGlyAsnAspIleuThrPheGlnHisIleArgAs 512
1647 TACCGATGAAGGGCGATGATTCACACCAATCAAGACAAGAATCA 1696
|||||
512 nValAspGlnGlyAlaArgIleValAsnHisAsnThrSerHisAlaSerT 529
1697 CCGTTACCATTCACAGCAATTAAGATTAATCT 1728
|||||
529 hTleThrIleuThrGlySerIleuThrAsnProAsnSerIleuSer 545
1728 1728
546 ValHisSerIleGlnAsnAspTyArgAspGlnAspAspTySerTyTyTy 562
1728 1728
562 ArgProArgArgProIleProGlnGlyLysAspLeuTyTyTyLysAsnT 579
1728 1728
579 yArgTyTyTyAlaLeuLysSerGlyGlySerValAsnAlaProMetPro 595
1729 1743
596 GlnAsnGlyValThrGlnAsnAsnAspTyrValPheMetGlyTyThrGln 612
1744 1768
612 nGlnGlnAlaLysLysAsnAlaMetAsnHisLysAsnAsnGlnArgIleS 629
1769 TTGCTTCACAGCGTGGTGGTGGAGAGAAAGTCAACCAAGCAAGCGG 1818
|||||
629 ergIlePheSerGlyPhePheGlyGlnGlnAsnGlyLysGlyHisAsnGly 645
1819 CGGCTCATGTGAATTAACCAACCGGAAGAACGGATGCGACTTACTGCT 1868
|||||
646 AlaLeuAsnIleuAsnPheAsnGlyLysSerAlaGlnAsnArgPheLeu 662
1869 TTCGGGCGGACAAATTTAAACGGCAATATCACGGCAAAACAGCGCAAC 1918
|||||
662 uThrGlyGlyThrAsnIleuAsnGlyLysIleSerValThrGlnGlyAsnV 679
1919 TGTTTTTCACCGGACAGCGACACCGCGCTTACATCATTTAGCAAGC 1968
|||||
679 alIleuLeuSerGlyArgProThrProHisAlaArgAspPheValAsnLys 695
1969 GGGTGTCAAAAATGGAAGT 2012
|||||
696 SerSerAlaArgLysAspAlaHisPheSerLysAsnAsnGlnValValRh 712
2013 GGACACGATGATGACGGGACCATTTAAAGCGGAAACTTCATAT 2061
|||||
712 eGlnAspAspTyrPheAsnArgThrPheLysAlaIleGlnIleAlaValA 729
2062 2109
729 snGlnSerAlaSerPheSerSerGlyArgAsnValSerHisIleThrAla 745
2110 GATTCGATTTAAGCAATCACGCCACAGCATTTTCGCTGCGCACCGCA 2159
|||||
746 AsnIleThrAlaThrAspAsnAlaLysValAsnLeuGlyTy 760
2160 TCAAAGCCACACATCTGTACACGTTCCGACGCGGTCTGACAAAGTT 2209
|||||
760 saAsnGlyAspGlnValLysValArgSerAspTyTyThrGlyTyValThrC 777
2210 GTACCGAAAAAACCATTAACCGAGTAAAGTGTGCTTCATTTAGACCAAG 2259
|||||
777 yAsnThrGlyAsnLeuSer 2309
2260 ACCGACATCAGAGGCAATGTACGCTTGGCGATCACGCTCATTTAAATCT 2309
|||||
793 ThrGlnIleAsnGlyAsnValAsnIleuAsnGlnAsnAlaIleValLe 809
2310 CACAGCACTGCCACACTCACCGGCAATCTTAGTGCAGCGGAGACAGCC 2359
|||||
809 u 820
2360 ACTATAGGTTACGCGCAACGCCACCAAAACGGCAACTCAGCTCGTG 2409
821 821
2410 GGCATGCCCAAGCAACATTTATACAGCCACATTAACGGCAACACATC 2459
|||||
822 GlyAsnSerArgValSerIleuAsnGln 830
2460 GGCTTCGACAAATGCTTCAATTAATCTAAGCAACAACGCCGTACAAACG 2509
830 830
2510 GCAGTCTGACGCTTCCGACAAACGCTAAGCAACGTAAGCCATTCGCCA 2559
830 830
2560 CTCACGCGCATGTCTCCCTAGCCGATTAAGCATTAATTCATTTGAAAA 2609
830 830
2610 CAGCGCTTTACCGAAAAATCAGCGGCGGCAAGGATACGGCATTAAC 2659
830 830
2660 TAAAGACAGCGATGACGCTGCCGTGGCGACGGAATTAAGCAATTA 2709
|||||
831 845
2710 AACCTTGACAAAGCCACCATTAACACTCAATCCGCTTGCACAGATGC 2759
|||||
846 SerLeuAlaAspSerHisIleHisLeuAsnAla 860
2760 GGCAGGCGCGCAACCGGACAGTGCAGATGCCGCCGCCCGCGCTTGC 2809
|||||
860 acIleSerAla 863
2810 GCCGCTCCCTATTAATCCGTTACGCCGCACTTGGCGGAATCCCGTTTC 2859
864 866
2860 AACAGCTGACGGTAAACGGAATTTGAACGCTCAGGGAACATTCGCGTT 2909
|||||
867 HisThrIleuLysIleAsn 882
2910 TATGTGGAACGCTTCCGCTACCGCGCAATTTGAAGTGGCGGAAA 2959
|||||
882 rIleuThrHisIleuAlaGlnAsnLeuGlyAspLysValLeuValLysGln 899
2960 GTTCGAGGACGCTTACACCTTGGCTGCAACATACCGGCAAGAACCC 3009
|||||
899 eArgIleSerGlyHisTyGlnIleuHisValGlnAspLysTyThrGlyGluPro 915
3010 GTAAGTCTGACGCAATGACGCTAGTGAAGAGAAAGCAACACACCGCT 3059
|||||
916 931
3060 GTCCGAAAATCTTAATTTACCTGCAAAAGCAACGCTGATGCCGCG 3109
|||||
```



```

4061 TCAGAAAGGCGATTACCGATACGAAAGCTCATATATCGCCACCCGCGC 4110
      :::::::::::::::::::: ||| ::::::::::::::::::::
1449 euSerSerAlaAspTyrLysLeuGlyAsnAspSerValLysValSer 1465
      :::::::::::::::::::: ||| ::::::::::::::::::::
4111 CTTCATTCACACCGCTACCGCGCGGCGCATTAAGCAGATTAATTCATCA 4160
      :::::::::::::::::::: ||| ::::::::::::::::::::
1466 MetSerValLysThrLeuThrAlaGlyLeuAspPheAlaTyrIraPhe 1482
      :::::::::::::::::::: ||| ::::::::::::::::::::
4161 ACCGGCGAACACATTCATCATCGCTTATTGAGCGCTGTCTAT...A 4207
      :::::::::::::::::::: ||| ::::::::::::::::::::
1482 s...ValGlyAsnLeuThrValLysProLeuLeuSerAlaAlaTyrPhe 1498
      :::::::::::::::::::: ||| ::::::::::::::::::::
4208 CCGATGCCGCTTCGGCAAGTCGCAACGGCGCATACCGCGCTTTG 4257
      :::::::::::::::::::: ||| ::::::::::::::::::::
1498 IAsnTyrGlyLysGlyValAlaSnValGlyLysAsnSerPheAlaTyr 1514
      :::::::::::::::::::: ||| ::::::::::::::::::::
4258 GCGGAGATTCGGCAAAACCCGAGTCGGGATGGGGCGTAACGCCGA 4307
      :::::::::::::::::::: ||| ::::::::::::::::::::
1515 LysAlaAspAsnGlnGlnIntYrSerAla.....GlyAlaAlaLeuLe 1529
      :::::::::::::::::::: ||| ::::::::::::::::::::
4308 AATCAAGGTTTCACGCTGCTCCACCGCTGCCCGCCGCAAGGGCGCC 4357
      :::::::::::::::::::: ||| ::::::::::::::::::::
1529 uTyrArgAsnValThrLeuAsnValAsnGlySerIleThrLysGlyLys 1546
      :::::::::::::::::::: ||| ::::::::::::::::::::
4358 AATTGAGAGCGCAGCAGCGCGGCGCATCAATTAGGCTACCGCTGG 4404
      :::::::::::::::::::: ||| ::::::::::::::::::::
1546 IntLeuGlnLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1561

```

seq_name: sp_bacteriap:09JXK3

seq_documentation_block:

ID 09JXK3 PRELIMINARY; PRT: 1431 AA.

```

DT 01-OCT-2000 (TREMBLEL 15, Created)
DT 01-OCT-2000 (TREMBLEL 15, Last sequence update)
DT 01-DEC-2001 (TREMBLEL 19, Last annotation update)
DE SRINE-TYPE PEPTIDASE.
GN NMB1998.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID:491;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-MC58 / SEROGROUP B:
RX MEDLINE=20175755; Pubmed=10710307;
RA Mettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Kechum K.A., Hood D.W., Feden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citlone H., Clark E.B.,
RA Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamathavan J.,
RA Gill J., Scariato V., Masiangni V., Pizsa M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL: AE002549; AAF42325.1;
DR MEROPS: S06.0PW;
DR TIGR: NMB1998;
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1.1.
DR PRINTS: PR00921; IGASERPTASE.
KW Complete proteome.
SQ SEQUENCE 1431 AA; 157645 MW; 6801F0E209C24ERC CMC64;

```

alignment_scores:

```

Quality: 1585.50 Length: 1668
Ratio: 1.829 Gaps: 50
Percent Similarity: 51.978 Percent Identity: 28.118

```

alignment_block:

US-09-303-518D-653 x 09JXK3 ..

Align seg 1/1 to: 09JXK3 from: 1 to: 1431

```

55 GGGCGGATCCGCTTCGCGCGCTTACCTTACCATATACCTGTCTGG 104
      :::::::::::::::::::: ||| ::::::::::::::::::::
17 GlyLeuPheAlaValSerProAlaTyrSerSerIleValArgAsn..... 31
      :::::::::::::::::::: ||| ::::::::::::::::::::
105 CATTCGCCCCCAAGCGCGCGGCGACACACTATTTCGGCATCACTAC 154
      :::::::::::::::::::: ||| ::::::::::::::::::::
32 .....AspAlaAspTyrG 36
      :::::::::::::::::::: ||| ::::::::::::::::::::
155 AATATATCGCACTTGGCGGAAATAAAGCAAGTTGCACTGCGGGCG 204
      :::::::::::::::::::: ||| ::::::::::::::::::::
36 IntYrPheArgAspPheAlaGlnAsnLysGlyAlaPheThrValGlyAla 52
      :::::::::::::::::::: ||| ::::::::::::::::::::
205 AAGATATTGAGGTTTACACAAAGGAGGAGTTGTCGCGCAATTCAT 254
      :::::::::::::::::::: ||| ::::::::::::::::::::
53 SerAsnIleSerIleGlnAspLysGlnLysIleLeuGlyArgValLe 69
      :::::::::::::::::::: ||| ::::::::::::::::::::
255 GACGAAGCCCGATGATTTTCTGTGTATCGCT...AACGGCG 301
      :::::::::::::::::::: ||| ::::::::::::::::::::
69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86
      :::::::::::::::::::: ||| ::::::::::::::::::::
302 TGGCGGCAATGGCGGGCGATCAATATATTGTGAGCGTGCGACATAACGC 351
      :::::::::::::::::::: ||| ::::::::::::::::::::
86 IeAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102
      :::::::::::::::::::: ||| ::::::::::::::::::::
352 GCGTATTAACAATGTTGATTTGTGGGAGGAGCAATCCGATCAGCA 401
      :::::::::::::::::::: ||| ::::::::::::::::::::
103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGlnGln 119
      :::::::::::::::::::: ||| ::::::::::::::::::::
402 CCGCTTTCTTACCAATTTGAAAAAGAAATATTATAAGCAGGAGCTA 451
      :::::::::::::::::::: ||| ::::::::::::::::::::
119 nAlaTyrThrTyrArgLeuValSerArg.....A 129
      :::::::::::::::::::: ||| ::::::::::::::::::::
452 ACCGGCATCCTTATGGCGCGGATTAATCATATGCCGCTTGCACAATT 501
      :::::::::::::::::::: ||| ::::::::::::::::::::
129 snProHisPro...AspTyrAspTyrHisLeuProAlaGlnAsnLysLeu 144
      :::::::::::::::::::: ||| ::::::::::::::::::::
502 GTACAGATGACAGACCTGTGTGATGACCACT..... 534
      :::::::::::::::::::: ||| ::::::::::::::::::::
145 ValThrGlnIleSerProThrAlaLeuSerSerValProLeuLeuGlyAs 161
      :::::::::::::::::::: ||| ::::::::::::::::::::
535 ....TATATGATGGGTGGAATACGCTGATTTAATAATACCTGATC 580
      :::::::::::::::::::: ||| ::::::::::::::::::::
161 nGlyGlnProLysAlaAsnAlaTyrLeuAspThrIraPheProTyrP 178
      :::::::::::::::::::: ||| ::::::::::::::::::::
581 GTGTTGGAATCGAGCAGCAGCAGCATATTGGCGGTCGATGAAGACGA 630
      :::::::::::::::::::: ||| ::::::::::::::::::::
178 heValArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGly... 193
      :::::::::::::::::::: ||| ::::::::::::::::::::
631 CCCAATAACCGCGAAGTTCATTCATTTGCAAGCGCATATTCTTGCT 680
      :::::::::::::::::::: ||| ::::::::::::::::::::
194 .....ThrArgThrArgThrAlaProAlaTyrGlnTyrLe 205
      :::::::::::::::::::: ||| ::::::::::::::::::::
681 CGTGGTGGAATACCTTGCACAAATGAGATCAGGTGCGACAGTCA 730
      :::::::::::::::::::: ||| ::::::::::::::::::::
205 uThrGlyGlyThrProLeuLysValLeuGlyPheGlnAsnHisGlyLeu 222
      :::::::::::::::::::: ||| ::::::::::::::::::::
731 ACTTAGTAGCGAATAATTAACATAGCCCATATGGTTTTTACCAACA 780
      :::::::::::::::::::: ||| ::::::::::::::::::::
222 euValArgGly...SerLeuThrAspGlnPro.....LeuAsnThr 234
      :::::::::::::::::::: ||| ::::::::::::::::::::
781 GAGGGTCATTTGGCAGACAGTGCACCAATGTTATGATGATGACCA 830
      :::::::::::::::::::: ||| ::::::::::::::::::::
235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLys 251
      :::::::::::::::::::: ||| ::::::::::::::::::::
831 AAAGCAAAAGTGTATTAATGGGTATTTGCAACAGCGCAACCCCTATA 880
      :::::::::::::::::::: ||| ::::::::::::::::::::
251 sGluAsnArgTyrPValLeuAlaGlyValLeuSerThr.....TyrA 265
      :::::::::::::::::::: ||| ::::::::::::::::::::
881 TAGGAAAAAGCAATGGCTTCAG.....CTAGTGTGTAAGATTGTTTC 924

```



```

265 1aglyPheaspasnPheaspasnLysTyrIleValThrGlnProGluPhe 281
295 TATGATGAATACTTTGGCTGAGATACCATTCAGTATTCTAGAACCA 974
282 ILeArgSerThrIleArgGln.....TyrGlnThrArg 292
975 T.....CAAAATGGAAATCTTTTAAAGCAAAATA 1009
292 glenAspValGlyLeuThrThrAsnGluLeuIleTrrPalgAspAsnGlyA 309
1010 ATGGCCAGAGAAATCGATGCCAAACATAAACATATTCTCTACCTTAT 1059
309 snGlyAsnSerThrLeuGlnGlyLeuAsnGluArgIleThrLeuPro... 324
1060 AGATTAAAAACAGAACCGTTCATTTGTTTATGTTTCTTATCCGAGAC 1109
325 .....IleAlaAsnProSerLeuAla..... 331
1110 AGCAGAGAACCTGTTTATCATGCTGCAGGTGGGTCAACAGTTATGCAC 1159
332 .....P 332
1160 CCAGACTGAATAATGAGAAAAATTTCTTATTGACAAAGAAAA... 1206
332 roGln...AsnAspSerArgHisMetProSerGluAspValGlySerThr 347
1207 .....GGTGAATTGATCTTACAGCA 1229
348 LeuIleLeuSerSerArgPheaspasnLysThrLeuMetLeuAlaAspAs 364
1230 CATCAACCAAGGCGGGCGGCTTTGTTTATTTGAGGTAATTTCAGGCT 1279
364 nIleAsnGlnGlyAlaGlyAlaLeuGlnPheaspSerAsnPheThrVal 381
1280 CGCCTAAACACACGAAGAGTGGCAAGCGGCGGCTTCATATCAGTGAT 1329
381 aIGlyLysAsnHis...ThrTrpGlnGlyAlaGlyValIleValAlaAsp 396
1330 GGCAGTACCGTACTTGAAATGAACGCGTGGCAACGACGCGCTGTC 1379
397 GlyLysArgValPheTrpGlnValSerAsnProLysGlyAspArgLeuSe 413
1380 CAAATCGCAAGACGACGCTGCTGTCACGCAACGAGGCAAAACCAAG 1429
413 rLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnGlyIleAsnGln 430
1430 GCTCGGTACGCTGGGCGACGCTAAAGTCATCTTAGATCAGACGCGGAC 1479
430 LysPheSerIleGlyGlnGlyThrValIleValLeuAlaGlnLysAlaAla 446
1480 GATCAAGGCAAAACCAAGCCTTTAGTAAATCGGCTTGTCAGCGGAG 1529
447 SerAspGlySerLysGlnAlaPheAsnGlnValGlyIleThrSerGlyArg 463
1530 GGGAGCGGTGCACTGAATGCCGATTAATCAGTTCAACCCGACAAACT 1579
463 gGlyThrIleValIleuAlaAspSerGlnGlnIleLysProGluAsnLeuT 480
1580 ATTGGGCTTTCGGGCGAGCTTGATGATTAAGAGGCGCATTCGTTTCG 1629
480 yTrpGlyPheArgGlyLeuArgLeuAspLeuAsnGlyLysAsnIleuAla 496
1630 TTCACCGCATTCAAATACCGATGAAGGCGATGATGTCACACAA 1679
497 PheThrHisIleArgHisAlaAspGlyLeuAlaGlnIleValAsnHisAs 513
1680 TCAAGCAAAAGATCCACCGTTACCATTAACGCAATTAAGATTTACTA 1729
513 nProAspGlnAlaIleThrLeuThrLeuThrGlyLysnProValLeuSerP 530
1730 CAACC..... 1734
530 roGluHisValGluTrpValGlnTrpGlyAsnAlaGProGlnGlyAsnAla 546
1734 ..... 1734
547 AlaValTyrGlyTyrIleAsnProHisArgAsnArgArgThrAspTyrPh 563
1735 .....GGCAAT..... 1740
563 eIleLeuLysProGlyGlyAsnProArgGluPhePheProLeuAsnMetL 580
1741 .....AACACAACTTGATAGCAAA 1761
580 yAsnSerThrSerTrpGlnPheIleGlyAsnAsnArgGlnGlnAlaAla 596
1762 AAAGAAATGGCC.....TACAA 1778
597 GluGlnValAlaGlnAlaGluAsnAlaArgProAspLeuIleThrPheG 613
1779 CGGTTGGTTGGCGAGAAAGATGCACAC..... 1806
613 yGlyTyrLeuGlyGluAsnAlaGlnThrGlyLysAlaAlaProSerTyrS 630
1806 ..... 1806
630 eLysThrAsnGluAlaAlaIleGluLysThrArgHisIleAlaAsnAla 646
1807 .....AAACGACGCGGCGCTCAATTCGAA 1832
647 AlaValTyrGlyArgProGluTyrArgTyrAsnGlyAlaLeuAsnLeuH 663
1833 TTACCAACCGGAGAGAGGATCGCACTTACCTTTCCGCGGACAA 1882
663 sTyrArgProLysArgThrAspSerThrLeuLeuLeuAsnGlyGlyMetA 680
1883 ATTTAAACGGCAATATCAGCGCAAAACGAGCAAACTGTTTTCAGCGGC 1932
680 snLeuAsnGlyGluValLeuIleGluGlyGlyAsnMetIleValSerGly 696
1933 AGACGACACCGCGCGCTTACATCATTTAGAGACGCGGTGTCAAAAT 1982
697 ArgProValProHisAlaTyrAspHisGlnAla..... 707
1983 GGAAGTATCCCAACAGAGAAATCGTGGAGCAACAGATTGGATGCACC 2032
708 .....LysArgGluProValLeuGluAsnGluTrpThrAspG 720
2033 GCACATTTAAAGCGGAAACTTCCATATTCAGGCG...GGACAAAGCGTG 2079
720 LyserPheLysAlaAlaArgPheThrLeuArgAsnHisAlaArgLeuThr 736
2080 GTTCCCGCAATGTTGCCAAAGTGGAAAGCGCAT.....TGCGATT 2120
737 AlaIleLysArgsnThrAlaHisLeuAspGlyAspIleThrAlaTyrAspLe 753
2121 AAGCAAT...CAGGCCCAAGCAGTTTCGGTGGCGACGACGATCAAGCC 2167
753 userGlyIleAspLeuGlyPheThrGlnGlyLysThrProGln..... 767
2168 ACACAAATCTTACACGTTGCGACTGACGGGTGTGCAAACTGTACCGAA 2217
768 .....CysTyrArgSerTyrHisSerGlySerThrHisCysThrPro 781
2218 AAACCATTTACCGACGATTAAGTGAATGATTGCTTCTTAGAGCAAGACGACAT 2267
782 AsnAlaValIleuLysAlaGluAsnTyrArgAlaLeuProAlaThrGlnVa 798
2268 CAGAGGCAATGTCAGCCTTGCGATCAGGTCATTTAAATCTCAGAGGAC 2317
798 ArgGlyAspIleThrLeuAsnAspArgSerGluLeuArgLeu..... 812
2318 TTGCGCACACTCAACGGCAATCTTAGTGACGGCGGACAGACGCACTTACG 2367
813 .....GlyLysAlaHis..... 816

```

```

2368 GTTACGGCAAGCGCCAAAGCGCAACCTCAGCCTGTGGCAATGC 2417
816 ..... 816
2418 CCAGCAACATTTAATCAAGCCACATTAAGCGACACATCGGCTTCG 2467
816 ..... 816
2468 ACAATGCTTCATTATCTAAGCAACAAGCCGTACAAAAGCGCAGTCG 2517
816 ..... 816
2518 ACGTTTCGCAACGCTAAGCAACGTAAGCATTCGCGACTCAACGG 2567
816 ..... 816
2568 CAATGTCCTCCAGCGATTAAGCAGTAATTCATTGAAAAAGCGCGCT 2617
817 .....L 817
2618 TTACCGAAAAATCAGCGCGCAAGATACGCATTACACTTAAA... 2664
817 eutyrglyserlearglyaspthrAlaValArgmetgluAla 833
2665 GACAGCGAATGAGCGCTGCGTCGGGACGGAATTAGCGCAATTAAACCT 2714
834 aspsersantrpThrleuserglinserseristhrglyAlaleuthle 850
2715 TGACAAGCCACCATTAATCAATTCGCTAATGACACGATGCGGCGAG 2764
850 uaspglylaglnIleThrleuasnProasppheAlasnasn..... 864
2765 GCGCGCAAAACGGCGAGTGGCGAGATGGCGCGCGCGCGCTTCGCGCGT 2814
864 ..... 864
2815 TCCTATTATCGTTACGCGCGCAACTTCGCGAGATCCGTTTCAACAC 2864
865 .....ThrlhsasnasnArgpheasnth 872
2865 GCTGACGGTAAACGCAAAATTGACGCTCAGGAGAACATTCGCTTATGT 2914
872 rleuthrValasnnglyThrleuaspglyPheglyThrpheargphelaut 889
2915 CGGAACCTGTGGCTACGCGAGCGGC...AATGGAACGTGGCGGAAGT 2961
889 hrGlyIleValArglysglnsnAlaProToloulysluengllyasp 905
2962 TCCGAAGCAGCTTACACCTTGCTGTCAACAATACCGGCAAGAACCGCT 3011
906 serArglyAlaPheglInIleHisValysasnThrlyglnIuProgl 922
3012 AAGTCTCAGCAATTGACGCTAGTGAAGAAAGACAACAACACCGCTGT 3061
922 nthrThrIguSerleuAlaLeuVal.....SerleuasnProIysshIs 937
3062 CGGAATCTTATATTTCACCCGTGCAAAAGCAACAGCATGCGCGCGCA 3111
937 erHisglInAlaArgphethrleuglnasnGlyTyrAlaAspleuglyAla 953
3112 TGGCGTTATCAGCTTATCCGCAAGAGCGGAGTTCGCGCTGCATATCC 3161
954 tyArgtyrIleLeuArglysnasnasnnglyTyrSerleuTyrasnPr 970
3162 GGTCAAAAGCAAGAGCTT..... 3180
970 oleuylsglnAlaIleuenglInlegluAlaThrArgAlaIuhisglua 987
3180 ..... 3180
987 rgAsnnglnAlaIatyrAsnnglnleuenglAlaThrAspIleSerArgln 1003
3181 .....TCCGCAAAACTCGGCAAGCGC..... 3201
1004 ValglInHisaspsersapAlaThrArgglInAlaIeuglnAlaItrpolnAs 1020
3202 .GGAGAAACAGAGCGCGCTTGACGGCAAAACAGACACACTTGGCGCA 3250
1020 nserglInThrgluLeuAlaArgIleaspsersglInValglInTyrleuserA 1037
3251 AACACAGCGCGGAAAAAGAC.....AACGGCGCA 3279
1037 laglnleuylsglnthrAspProleuthrGlyIleleuthrArgAlaIn 1053
3280 AGCCTTGACGCGCTGATTCGCGCGCGCGCAATGCC..... 3315
1054 Asnleucysala.....AlaInnglyTyrSerAlaAspIleCysArgsl 1068
3315 ..... 3315
1068 nValAlaIysAlaAlaAspThrAsnAspleuthrleuphegluThrGluL 1085
3316 .....ACGAAAGCAGAAAGTGTTGCCGACCG 3345
1085 euAspThrTyrIlegluArgValgluMetAlaIleuSerIleuAspIys 1101
3346 GCCCGCAGCGACGCGCGGAAAAATGCCGATTATGACGCGGAGGAGACA 3395
1102 AlaArglnnglygllyasp.....AlaInAlaIleValgluThr 1113
3396 GAAAAAAGCGGTGACGCGGATTAAGACACCGCTTGGGGAACAG...C 3442
1113 rAlaArgHisAlaIatyrleuasnAlaIeuasnArgleuserArgglInIleH 1130
3443 GCGAAGCGGAAACCGCGCGCTTACCAACCGCTTCCCGCGCGCGCGC 3492
1130 isserleuylstrnglyValAlaIleArgmetProasnleu..... 1144
3493 GCCCGCGGGATTGGCGCAACCGCACGCCCAACCGCAACCCCAACCGCA 3542
1144 ..... 1144
3543 GCGGACCTGATCAGCGCTTATGCAATAGCGGTTGAGTAATTTCG 3592
1145 .AlaIleuIleSerArgserAlaasnThrAlaValSerIleuglnAla 1161
1161 latyrasnThrIleArglnnglnAlaIleArgIleasphrghIsleu 1177
3593 CCACGCTCAACAGCGTTTTCGCGTACAGAGCAATGAGCGCGTGT 3642
1161 latyrasnThrIleArglnnglnAlaIleArgIleasphrghIsleu 1177
3643 GCGGAAGACCGCGCAACCGCGTTTGACACAGCGGACCGGACACCA 3692
1178 ThrAspProIleInnglnsnIletrleuglnuThrGlyThrIleInngln 1194
3693 ACACTACCGTTCGCAAGATTTCGCGCTACGCGCAACCAACGACCTGC 3742
1194 rAspTyrHisSerGlyThrHisArgProtyrIleInnglnThrAsnTyrA 1211
3743 GCCAATCGGTATCGAAAAAACCTCGGACGCGG...CGCGTGCATC 3789
1211 lanHisIleglIleInnglnThrGlyIleThrAspArgleuserValgluThr 1227
3790 CTGTTTTCGCAACCGGACCGGAAACACCTTGACAGAGCGGACATCGC 3837
1228 IleleuthrAspIleuArgIletrAsnAlaArgPheaspglnIleValSerAl 1244
3838 .....AATCGCACGCGCTTCCACGCGTGGCGCTTTCGGGCAATACGCA 3883
1244 aArgAsnArgSerAsnIleAlaHisIleuPheValIysglIleuasnclyA 1261
3884 TCGGCAAGTTTCGACATCGGCATCGCGCGCGCGGTTTATAGCGG 3933
1261 la.....LeuPheAlaIleAlaAspleuIleTyrSerAsnSer 1273
3934 .....AGCCTTCAGACGCGCATCAGAGCGCAAAATCCGCCCGCGTGTCT 3977

```

```

1274 ArgThrArgPheThrAspTyrAspGlyAlaAlaValArgArgHisAlaTr 1290
      :::::|||||
3978 GCATTAGCGGATTCAGCGCAAGATACCGCGGAGTTGGCGGATTCGGCA 4027
      |||||:::
1290 pAspAlaGlyIleAsnThrGlyIleIleAsp...ThGlyIleAsnL 1306
      :::::
4028 TCGAACCGCGCACATCGC.....GCACGCGGTAT 4056
      :::::|||||
1306 euArpProTyrAlaGlyIleArgIleAsnArgSerAsnGlyAsnArgTyr 1322
      :::::|||||
4057 TTGTCCTCAAAAGCGGATTCAGATACGAAACGTCATATTCGCCACCCC 4106
      :::::|||||
1323 ValIleuAspGlyAlaGlu.....IleAsnSerP 1332
      :::::
4107 GGGGCTTGCAATTCACCCGCTACCGCGCGGATTAAGCGATTAATTCAT 4156
      :::::|||||
1332 oAlaGlnIleGlnThrThrPHisAlaGlyIleArgIleuAspLysThrV 1349
      :::::|||||
4157 TCAACCGCGCGCACATTCCTCATCGCCTTATTTGAGCCCTGCTCAT 4206
      :::::|||||
1349 aGlnIleuGlyGln...AlaLysIleuThrProIlePheSerSerAspTyr 1364
      :::::|||||
4207 .....ACGATGCCGCTTCGGCGCAAGTCCGACCGCGGTCAATACCGC 4250
      :::::|||||
1365 TyrHisThrArgGlnAsnSerGlySerAlaLeuSer...ValAsnAspAr 1380
      :::::|||||
4251 CGTATTCGGCGCAGATTCGGCAAAACCGCGCATGCGGAATGGCGGTAA 4300
      :::::|||||
1380 gThrIleuGlnGlnAlaAlaHisGlyThrIleuHisThrIleuGlnIleA 1397
      :::::|||||
4301 ACCCGCAATCAAAAGTTTCACGCTCCCTCCACGCTCCCGCGCGCAAG 4350
      :::::|||||
1397 sPAlaGlyTyrGlyTyrAsnAlaLysIleuHisAlaAlaTyrGlyLys 1413
      :::::|||||
4351 GGGCGCGCAATTCGAGCGCAGCAGCAGCGGCGATCAATTAAGCTACCG 4400
      :::::|||||
1414 AspSerAsnThrAlaArgHisLysGlnAlaGlyIleLysIleGlyLys 1430
      :::::|||||
4401 CTGG 4404
      :::::
1430 nTrp 1431

```

seq_name: sp_bacteriap:Q9JVB9

seq_documentation_block:

ID Q9JVB9 PRELIMINARY; PRT; 1773 AA.

AC Q9JVB9;

DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DE IGA1 PROTEASE (EC 3.4.21.7).

GN IGA1 OR NMA0905.

OS Neisseria meningitidis (serogroup A).

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

ON NCBI_TaxID=65699;

RP SEQUENCE FROM N.A.

RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;

RX MEDLINE=20222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,

RA Klee S.R., Morelli G., Basham D., Brown K., Chillingworth T.,

RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,

RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,

RA Whitehead S., Spratt B.G., Barrall B.G.,

RT "Complete DNA sequence of a serogroup A strain of Neisseria

RT meningitidis 22491."

RL Nature 404:502-506(2000).

DR EMBL: AL162754; CAB84182.1;

DR MEROPS: S06.001; Dihydroorotase.

DR InterPro: IPR002195; Dihydroorotase.

DR InterPro: IPR000710; IGA_S6.

```

DR Pfam; PF02395; IGA1; 1
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease; Hydrolase; Complete proteome.
SQ SEQUENCE 1773 AA; 196350 MW; CAC19E7313D76CE1 CMC64;

```

alignment_scores:

Quality: 1515.00 Length: 1958
Ratio: 1.647 Gaps: 53
Percent Similarity: 46.987 Percent Identity: 24.974

alignment_block:

US-09-303-518d-653 x Q9JVB9

Align seg 1/1 to: Q9JVB9 from: 1 to: 1773

```

49 AAACCGCGCGCGATCGCTTCGCGCCGCTTACTTACCATATGCTGTC 98
||||| :::::
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
:::
99 GTTCGGCATTCGCCCCAAGCCCGCGCGGACACACTTATTCGGCATCA 148
:::
18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgSpAspValA 35
:::
149 ACTACCATACTATGCGCATTCGCGCAAAATTAAGCAAGTTCGACATC 198
:::
35 sPTrpGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
:::
199 GGGCGGAAGATATGAGGTTTACACAAAAAGGGGATGGCGGCA 248
||||| :::::
52 GAlaThrAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAs 68
:::
249 ATCGATGACGAAAGCCCGCATGATGATTTTCGTGGTATCGGTAC 297
:::
68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspAlaAsnA 85
:::
298 ..GGCGTGGCGCATTCGCGCGCGCATCAATATTCGCGCGCAT 345
:::
85 rGArgThrLeuThrValIleAsnProGlnTyrAlaValSerValLysHis 101
:::
346 AACGCGGC.....TATACAAATGTTGTTGGTGGCGA 380
|||
102 ValLysGlyAspGluIleSerTyrGlyHisHisAsnGlyHisLeuAs 118
:::
381 GGAAGCATATCCGATCAGCACCGCTTTCTTACCAATTTGTAAGAA 430
:::
118 pValSerAsnAspGluAsn.....GluTyrArgSerValAlaGlnA 132
:::
431 ATAAATTAAGACGAGGACCTAACGCGCATTCCTTATGGCGGC 471
||||| :::::
132 sNAspTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArg 147
:::
472 .....GATTCATCATATGCGCGCTTTCACAAATTTGTCAAGATGAGA 515
||||| :::::
148 LeuGluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAl 164
:::
516 ACCGTGAGATGACCGAGTTATATGATGGGTGGA...TACGCGATT 562
||||| :::::
164 aProIleAlaProThrSerAlaGlyGlyValGluThrTyrLysAspL 181
:::
563 TAAATTAATACCTGATCGTTCGATTCGACAGCAGCAGCAATATGG 612
||||| :::::
181 yAsnArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe... 196
:::
613 CGGCTGATGTAAGACGACCATTAACCGGAAAGTTCATATCATATTCG 662
||| :::::
197 .....GluTyrAsnSerArgTyrAsnMetThrGluLeuSe 208
:::
663 AAGGCATATTCCTTGCTCGTGGCATTAACCTT..... 699
:::
208 rArgAlaTyrArgTyrAlaIleAlaIleGlyThrProTyrGlnAspValAsnV 225

```


783 ..AspLysAlaLeuAsnSerPheGlyAlaThrLysIleAsnGlyAsnVal 798
2281 AGCCTTGGCGATCAGCTCATTTAAATCTCAACGACTTGGCACACTCA 2330
:::||||| ::||| ||| ||| ::|||
799 AsnLeuAsnGlnAsnAlaIleValLeu...GlyLysAlaAlaIleThr 814
2331 CGGCATCTTGAATGCGAGCGGAGACACGACTATACGTTACGGCAACG 2380
|||:::|||||
814 pGlyGlnIleGlnGly..... 819
2381 CCACCCAAAGCGCAACCTCAGCCTCGTGGCAATGCCCAAGCAACATTT 2430
|||:::|||||
820GlnGlyAsnSerArgValSerLeu 827
2431 AATCAAGCCACATTAACGCGACACATCGGCTTCGGACATGCTTCATT 2480
|||::|
828 AsnGln..... 829
2481 TAACTTAAGCAACAACGCCGTACAAAAGCGCATGTGACGCTTCCGACA 2530
829 829
2531 ACGCTAAGGCAACGTAAGCCATTCCGACCTCAACGGCAATGCTCCCTA 2580
829 829
2581 GCCGATAAGGCATGATTCCATTTTGAAGAAGCGGCTTTACCGGAAAT 2630
829 829
2631 CAGGGGGGCGAGATAGCGCATTTACACTTAAAGACAGCAATGACGCG 2680
|||:::|||||
830HisSerLysTrpHisL 835
2681 TGGCGTGGGACGCAATTAAGCAATTTAAACCTTGACACGGCACCATTT 2730
|| ::|:::|||||
835 euThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIle 851
2731 ACACTCAATTCGCGCTATCGACACGATGCGGACGGCGCAACCGGACG 2780
|||:::|||||
852 HisLeuAsnAsnAla...SerAspAlaGlnSerAla..... 862
2781 TGGCGGAGATGCGCGCGCGGCTTCCGCTTCCCTATTTCGCTTA 2830
862 862
2831 CGCGCCCACTCGCAGAAATCCCGTTTCAACAGCGTGAAGTAAACGCG 2880
863AsnGlnTrpHisThrLeuLysIleAsn... 871
2881 AAATTGACGCGTCAAGGACATTCGCTTATGTGGAACCTTTCGCTA 2930
|||:::|||||
872 HisLeuSerGlyAsnGlnHisPheHisTyLeuThrHisLeuAlaGlnAs 888
2931 CCGCAGCGGCAATTAAGCTGGCGGAAGTTCCGAAGCAGCTTAACCT 2980
::|::|:::|
888 nLeuGlyAspLysValLeuValLysSerAlaSerGlnHisTyGlnL 905
2981 TGGCTGTCAACAATACCGGACAGCAACCGTAACTTCGACGAATGAGG 3030
|||:::|||||
905 euHisValGlnAspLysThrGlyLysPro...AsnGlnIleGlyLeuAsn 920
3031 GTAGTGGAAAGAAAGACACACACCGCTGCGAAATCTTAATTTTCAC 3080
:::|:::|||||
921 LeuPheAspAlaSerSerValArgAspArgSer...HisLeuSerValS 936
3081 CTTGCAAAAGCAACGCTGATCGCGGCGGCTTATCACTTATTC 3130
||| |||:::|||||
936 rLeuAlaAsnHisHisValAspLeuGlyAlaLeuArgTyThrIleLysT 953
3131 GCAAGACGCGCAGTTCCGCTGCATTAATCCG..... 3162
:::|||||
953 hGlnAsnGlnLysIleThrArgLeuTyAsnProTyAlaGlnAsnArgArg 969
3162 3162
970 ArgValLysProAlaProSerProAlaThrAsnThrAlaSerGlnAlaG 986
3162 3162
986 nThrAspSerAlaGlnIleAlaLysProGlnAsnIleValAlaProp 1003
3162 3162
1003 roSerProGlnAlaAsnGlnAlaGlnAlaLysArgGlnGlnAlaLys 1019
3163CTCAAGACACA..... 3174
1020 AlaGlnGlnValLysArgGlnGlnAlaGlnAlaGlnLysSerAlaG 1036
|||:::|||||
3174 3174
1036 uLeuAlaLysGlnLysAlaGlnAlaGlnLysArgGlnAlaArgGlnLeuAlaT 1053
3174 3174
1053 hArgGlnLysAlaGlnGlnLysArgSerSerAlaGlnLeuAlaArgArg 1069
3175GAGCTTCCGACAAA..... 3189
1070 HisGlnLysGlnArgGlnAlaAlaGlnLeuSerAlaLysGlnLysValG 1086
|||:::|||||
3189 3189
1086 uAlaGlnArgGlnAlaGlnAlaLeuAlaValArgArgLysAlaGlnAlaG 1103
3190CTCGGCAAGCGGCGGCAACAG 3213
1103 GlnGlnAlaLysArgGlnAlaAlaGlnLeuAlaArgArgHisGlnLysGln 1119
3214GCCGCTTGACGGCAAAACAG..... 3234
1120 ArgGlnAlaAlaGlnLeuSerAlaLysGlnArgValGlyGlnGlnLys 1136
|||:::|||||
3234 3234
1136 gArgGlnThrAlaGlnSerGlnProGlnArgArgLysArgAlaAlaP 1153
3234 3234
1153 roGlnAspTyrMetAlaAlaSerGlnAspArgProLysArgArgGlyHis 1169
3234 3234
1170 ArgSerValGlnGlnAsnAsnValGlnIleAlaGlnAlaGlnLys 1186
3235GCACAACTTGCAGCAAC 3253
1186 uAlaArgArgGlnGlnGlnLysArgLysAlaAlaGlnLeuLeuAlaLysG 1203
3254 AACAGCGGGAAGAAAGAC...AACGGCAAGCCTTGACGCGCTG..... 3294
|||:::|||||
1203 hArgAlaGlnAlaGlnLysArgGlnAlaGlnAlaLeuAlaArgArgLys 1219
3294 3294
1220 AlaGlnAlaGlnGlnAlaLysArgGlnAlaAlaGlnLeuAlaHisArgGln 1236
3295ATTGCGCGCGGCGGCAATGCCACG 3319
1236 nGlnAlaGlnArgLysAlaAlaGlnLysSerAlaAsnGlnLysAlaAla 1253
3320 AAAAGCGCAAGAAGTGTGCC..... 3339
1253 lArgAlaGlnAlaLeuAlaAlaArgGlnGlnLysAlaLeuAlaArgGln 1269

CC BY-NC-ND 4.0 International license.

```

435 lYValaInGlnGlyInLeuLysValGlyAspGlyThrValIleLeuAsn 451
1468 CAGCAGCGCGGATGATCAAGCAAAAAAAGCCTTAGTGAATCGCCTT 1517
|||||
452 GlnGlnAlaAspAlaAspLysValGlnAlaPheSerGlnValGlyIle 468
1518 GGTGACGGGCGGCGGACGGTGCAGCAATGCCGATATCAGTCAACC 1567
|||||
468 eValSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnLeuAsn 485
1568 CCGACAAACTCTATTTCGGCTTCGCGCGAGCTTTGATTGAAACGGG 1617
|||||
485 roAspAsnLeuThrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGly 501
1618 CATTGCGCTTTGTTCCACCGCATTCAAAATACCATGAGGGCGGATGAT 1667
|||||
502 AsnAspLeuThrPheGlnHisIleArgAsnValAspGlnGlyAlaArgIle 518
1668 TGTCAACCATCATCAAGCAAAAGATCCAGCTTACCATTCAGCAGCATA 1717
|||||
518 eValAsnHisAsnThrGlyHisAlaSerThrIleThrLeuThrGlyLys 535
1718 AAGATATTACT..... 1728
535 erLeuIleThrAspProLysThrIleSerIleHisTyrIleGlnAsnAsn 551
1728 ..... 1728
552 AspAspAspAlaGlyTyrTyrTyrTyrArgProArgLysProIlePr 568
1728 ..... 1728
568 oGlnGlyLysAspLeuTyrPheLysAsnTyrArgTyrAlaLeuLys 585
1729 ..... ACAACGGGCAAT 1740
585 erGlyGlySerValAsnAlaProMetProGlnAsnGlyGlnThrGlnAsn 601
1741 AAC..... 1743
602 AsnAspTrpIleLeuMetGlySerThrGlnGlnGlnAlaLysLysAsnAl 618
1744 ...ACAACCTTGATAGCAAAAAAAGAAATTGCTACACGGTGGTTTG 1789
|||||
618 aMetAsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPhePheG 635
1790 CGGAAAGATGCAACCAAAAGACGGCGGCTCAATCGAATGAAATCCAA 1839
|||||
635 lYglnGlnAsnGlnLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsn 651
1840 CCGGAAGAAGCGGATCGCATTACTGCTTCGCGGAGCAAAATTTAA 1889
|||||
652 GlYLySerAlaGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAs 668
1890 CGGCATATTCAGCAACAAACGCAAACTGTTTTCAGCGGAGACCGGA 1939
|||||
668 nGlyLysIleSerValThrGlnGlyAsnValLeuLeuSerGlyArgProT 685
1940 CACCGCAGCCCTCAATCATTTAGAGAGCGGCTGTCAAAATGGAAGT 1989
|||||
685 hPProHisIleAlaArgAspPheValAsnLysSerSerAlaArgLysAspAla 701
1990 .....ATCCCAAGAGAGAAATCGTGGGAGCAACGATTTGGATCGACG 2033
|||||
702 HisPheSerLysAsnAsnGlnValAlaPheGlnAspAspTrpIleAsnAr 718
2034 CACATTTAAAGCGGAACCTTCATATT...CAGGGCGGACAAAGCGGTGG 2080
|||||
718 gThrPheLysAlaThrGlnIleAlaValAsnGlnSerAlaSerPheSers 735
2081 TTTCCCGCAATGTTGCCAAAGTGGAAGCGGATTTGCATTAGCATATCAC 2130
|||||
735 erGlyArgAsnValSerAspIleThrAlaAsnIleThrAlaThrAspAsn 751
2131 GCCCAAGCGATTTCGGTGTGCGACCGCATCAAAAGCCACAAATCTGAC 2180
|||||
752 AlaLysValAsnLeuGlyTyr....LysAsnGlyAspGlnValCysVal 766
2181 ACCTTCGCACTGGACGGGTGTGACAAAGTTGTACGAAAAAACCATTTACC 2230
|||||
766 lArgSerAspTyrThrGlyTyrValThrCysAsnThrGlyLysLeuSer 782
2231 ACGATTAATGATGTTGTTTATGACCAAGACCGACATCGAGCAATGTC 2280
|||||
783 ..AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyAsnVal 798
2281 AGCCTTGGCGATCACGCTCATTTAAATCTACAGAGCATTTGCCACATGCA 2330
|||||
799 AsnLeuAsnGlnAsnAlaIleValLeu...GlyLysAlaAlaLeuThr 814
2331 CGGCAATCTTAAGTCAGCGCGGAGACACGCATATACGGTTACGGCGAACG 2380
|||||
814 pGlyGlnIleGlnGly..... 819
2381 CCACCAAAAAGCGCACACCTCAGCGCTGTGGCAATGCCACACATTT 2430
|||||
820 .....GlnGlyAsnSerArgValSerLeu 827
2431 AATCAAGCCACATTAACGCAACACATCGGCTTGAGACATGTTCAAT 2480
|||||
828 AsnGln..... 829
2481 TAATCTAAGCACACACCGCGTCAAAAGCGAGCTGACGCTTTCGGACA 2530
829 ..... 829
2531 ACGCTAAGCGCAACGTAAAGCATTTCCGACATCAACGCAATGTCTCCSTA 2580
829 ..... 829
2581 GCCGATTAAGCGAGTATTCATTGTAAGACGCCGCTTACCGGAAAT 2630
829 ..... 829
2631 CAGCGGCGGCAAGATACGSCATTACACTTAAAGACAGCAATGGAGCG 2680
|||||
830 .....HisSerLysTrpHisLe 835
2681 TGCGGTGCGGCGAGCAATTAAGCAATTTAAACCTTGACAAAGCCACATTT 2730
|||||
835 euThrGlyAspSerGlnValHisPheHisLysLeuAlaAspSerHisIle 851
2731 ACACSTCAATTCCGCTATGTCGACACGATGCGGCGAGCGCGCAAAACGGCAG 2780
|||||
852 HisLeuAsnAsnAla....SerAspAlaGlnSerAla..... 862
2781 TGCGGCGAGTGGCGCGCGCGCTTCGCGCTCCCTATATCCGTTA 2830
862 ..... 862
2831 CGCGCGCAACTCGGAGCAATCCCGTTTCAACAGCTGACGCTAAACGGC 2880
|||||
863 .....AsnLysTyrHisThrLeuLysIleAsn... 871
2881 AAATTGAACGTACAGGACATTTCCGTTATATGTCGGAATCTTCGGCTA 2930
|||||
872 HisLeuSerGlyAsnGlnHisPheHisTyrLeuThrHisLeuAlaLysAs 888
2931 CCGCAGCGGCAAAATTGAAGCTGGGCGGAAATTCGGAAGCAGCATTTACST 2980
|||||
888 nLeuGlyAspLysValLeuValLysGlnSerAlaSerGlyHisTyrGlnL 905
2981 TGCGTGTCAACAAATACCGGACGCAACCGCTAAGTCTCGACAAATTCAGC 3030
|||||
905 euHisValGlnAspLysThrGlyLysPro...AsnGlnGlnGlyLeuAsn 920

```



```

1520 AlaAlaAlaGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSe 1536
3382 TGAATTTTCCGCCACGCTCAACACCGTTTCCGCC..GTACAGACCAAT 3628
1536 rAspAlaMetAlaSerThrGlnSerIleLeuLeuAspThrGlyAlaSerI 1553
3629 TGGACCGCGGTTTGGCCGAACCGCCGCC.....AACGCC 3663
1553 eutHrAghHisIleAlaGlnLysSerArgAlaAspAlaGlnLysAsnSer 1569
3664 GTTTGGACACAGCGGCATCCGGGACACCAACACTACCGTTCCGAAGATT 3713
1570 ValTrpMetSerAsnThrGlyTyrGlyArgAspTyrAlaSerAlaGlnTy 1586
3714 CCGGCGCCACCGCCCAACAAACCGACCTGCCCAATCGGTAATGGAGAAA 3763
1586 rArgArgPheSerSerLysArgThrGlnThrGlnIleGlyIleAspArgS 1603
3764 ACCTCGCGACGCGG...CGCGTCGGCATCTGTTTGGCACACACCGGACC 3810
1603 erLeuSerGlnAsnMetGlnIleGlyValLeuThrTyrSerAspSer 1619
3811 GGAACACCTTTCAGACGACGCGCATCGGCACACTCGCAGCGCTGCCACGG 3860
1620 GlnHsrThrPheAspGlnAlaSerGlyLysAsnThr...PheValGlnAl 1635
3861 TGGCGTTTGGGCAATACGCGCATCGCGC...AGGTGCGACATCGGCATCA 3907
1635 aAsnLeuTyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrValAlaGlyA 1652
3908 GCGCGCGCGCGGTTTGTAGACGCGGACCTTTTCAGACGCGCATCAGAGC 3957
1652 spLleGlyAlaGlySerLeuArgSerArgLeuGlnThrGlnGlnLysAla 1668
3958 AAAATCCGCCCGCGCTGCTGCATTACGCGCATTCAGGCAACATACCGCGC 4007
1669 AsnPheAsnArgAlaSerIleGlnThrGlyLeuThrLeuGlyAsnThrLe 1685
4008 AGGTTTCGGCGGATTCGGCATCGAACCGCATCGCGCACGCGCTATT 4057
1685 ulysIleAsnGlnPheGlnIleValProSerAlaGlyIleArgTyrSerA 1702
4058 TCGTCCAAAAGCGGATTACCGATACGAAACGTCATATTCGACACCCCG 4107
1702 rGleuSerSerAlaAspTyrLysLeuGlyAsnAspSerValLysValSer 1718
4108 GGCTTTCGATTTCACCGCTACCGCGCGCGCATTAAGGACATTTATTCATT 4157
1719 SerMetSerValLysThrLeuThrAlaGlyLeuAspPheAlaTyrArgPh 1735
4158 CAACACCGCGCAACATTTCCATCAGCGCTTATTTGAGCCTGTCCTAT.. 4206
1735 eLys...ValGlyAsnLeuThrValLysProLeuLeuSerAlaAlaTyrP 1751
4207 ..ACCGATGCCGCTTCGCGCAAAAGTCCGACGCGCGCTCAATACGCCGTA 4254
1751 heAlaAsnTyrGlyLysGlyValAlaAsnValGlyLysAsnSerPheVal 1767
4255 TTGGGCGCAGGATTTGGCAAAACCGCGCAGTCGGGATGGGCGGTAAAGC 4304
1768 TyrLysAlaAspAsnGlnGlnIleTyrSerAla.....GlyAlaAlaLe 1782
4305 CGAAATCAAGGTTTCACGCTGCCCTCCACGCTCCCGCGCCGCAAGGGGC 4354
1782 uLeuTyrArgAsnValThrLeuAsnValAlaAsnGlySerIleThrLysGlyL 1799
4355 CGCAATTTGACGACGACACGCGCGCGCATCAATTTAGCGTACCGCGTG 4404
1799 ySglnLeuGlnLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1815
seq_name: sp_bacteria:Q93T34
seq_documentation_block:

```

```

ID Q93T34 PRELIMINARY; PRT; 1764 AA.
AC Q93T34;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE IGA1 PROTEASE TYPE 2.
GN IGA1.
OS Haemophilus aegyptius.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=725;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F3031;
RA McGillivray G., Actis L.A.;
RT "Iga protease from H. aegyptius F3031."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369907; AAK56925.1; -.
KW Protease.
SQ SEQUENCE 1764 AA; 193863 MW; EC583CCDE81DBE6C CRC64;

```

```

alignment_scores:
  Quality: 1460.50      Length: 1933
  Ratio: 1.554          Gaps: 57
  Percent Similarity: 48.629  Percent Identity: 24.004

```

```

alignment_block:
US-09-303-518D-653 x Q93T34 ..

```

```

Align seg 1/1 to: Q93T34 from: 1 to: 1764

```

```

64 CGCTTCGCGCGCTTACTTAGCATATGCTGCTGCGCATTCCTGCG 113
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
5 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
114 CCAAGCCCGCGGCGGACACACTTATTTCGGCATCAACTCAATACTATAC 163
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
21 oTyrThrGlnAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38
164 GCGACTTTCGCGGAAATTAAGGCAAGTTTCAGTCGCGCGCAAGATATT 213
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
38 rGAspPheAlaGlnAsnLysGlyArgPheSerValGlyAlaThrArgAsnVal 54
214 GAGGTTTACAAACAAAAGGAGGAGTTGTCGCGCAAAATGATGACGAAAC 263
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
55 GlnValArgAspLysAsnAsnHisSerLeuGlyAsnAlaLeuProAsnGln 71
264 C...CCGATGATGATTTCGTGTGATTCGCGTAAAC...GGCGTGGCGG 307
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
71 yIleProMetIleAspPheSerValAlaAspValAlaAsnLysArgIleGlyT 88
308 CATTTGGCGCGCGATCATATATATTGTAGCGTGCACATACGCGGCTATP 357
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
88 hIleuValAspProGlnTyrIleValSerValLysHisAlaHisGlnTyr 104
358 AACAAATGTTGATTT.....GTGCGGAGGGAAGCAA 389
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
105 MetAsn...AspPheTyrPheGlnHisTyrAsnGlnHisArgAspValSe 120
390 TCCGATCAGCAGCCGCTTTCTTACCAAAATTCGAAAGAAATGATATATA 439
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
120 rAsnAspGlnAsnLys.....TyrSerValAlaThrGlnAsnAsnValAla 135
440 AAGCAGGAGACT.....AACGCCATCTTATGCGCGCGATTTATCATATG 483
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
135 snsrSerGlnLysTrpAspValAlaAsnLysArgLeuAspArgLysAsnMet 151
484 CCGCGTTTGCACAAATTTTTCACAGATGCGAAGACCTGTTGAGTGCACAG 533
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
152 ProArgLeuAsnLysPheValThrGlnValAlaProThrThrProThrLe 168
534 TTATATGATGGGTGGAAA...TACGCTGATTTAAATTAATACCGCATC 580

```

[illegible]

```

2069 GACACGCGGTGGTTC...CGCAATGTTGCCAAAGTGAAGCGGATTCG 2115
      :: ::::: ||| ||||| ||| ::::: :::::
750 snalSerLeuTyrSerGlyArgSerValGlySerIleThrSerAsnIle 766
2116 CATTAAAGCAATCAGCCCAAGCAGATTTTGGTGTGCACGCGCATCAAG 2165
      |||::: |||||::: :::::
767 ThrAlaSerSerLysIleGlnValHisIleGlyTyr.....LysLysG1 781
2166 CCACACATTCCTACAGCTTGGAGTGGAGCTGTGACAGTGTACCG 2215
      : |||::: ||||| ||||| ::::: |||
781 yAspThrValLysValArgSerAspTyrThrGlyValAlaThrCysHisA 798
2216 AAAAACCATTACCGCAGATTAAGTGAATGCTTCATTGAGCAACCGAC 2265
      ::::: ::::: ::::: |||::: |||:::
798 snGlyAspLeuSer...GluLysAlaLeuAsnSerPheAsnAlaThrAsn 813
2266 ATCAGAGCGAATGTACGCTTGGCGATCAGCGCTCATTTAAATCTCACAG 2315
      :: |||||::: |||::: ::|||::: |||
814 ValPheGlyAsnValAsnLeuSerGlyAsnAlaAsnPheThrLeu...G1 829
2316 ACTTGCCACCTCACAGCGCAATCTTAGTCAGCGCGAGACACGCACTATA 2365
      | |||||::: |||::: ||||| ||||| :::::
829 yLysAlaAsnLeuHisGlySerIleGlnAlaGlyLysAsnSerGlnValH 846
2366 CGGTTACCGCGCAACGCCCAACCGCAACCTCAGCCCTCGTGCGCAAT 2415
      ::|||::: |||
846 IAspThrGlnAsn..... 850
2416 GCCCAAGCAACATTTATCAAGCCACATTTAAACGGCAACACATCGGCTTC 2465
      850 ..... 850
2466 GGACATGCTTCATTATTAACTAAGCAACAGCGCTACAAACGGCACTC 2515
      850 ..... 850
2516 TGACGCTTCCGCAACGCTAAGGCAACGTAAAGCCATTCCGCACTCAAC 2565
      850 ..... 850
2566 GGCATGTCTCCCTAGCCGATAGGACAGATTTCCATTTGAAACAGCCG 2615
      850 ..... 850
2616 CTTTACCGGAAATCAGCGCGGCAAGATACGCGCATTAACCTTAAAG 2665
      850 ..... 850
2666 ACAGCGAATGACGCTGCCGTCGCGGACGCAATTAGGCAATTTAAACCTT 2715
      |||::: ||| ::::: ::|||::: |||
851 ..SerHisIleTyrLeuThrGlyAspSerAsnValHisGlnLeuAspLeu 866
2716 GACAAACGCGCAGCATTAACCTAATCCGCTTACGACACGATGCGGCAAG 2765
      |||::: ||| ||||| ||||| :::::
867 LysAsnGlyHisIleHisLeuAsnSerAlaAspAsnLysAsn..... 880
2766 CCGCGAAACCGGCAAGTGGCGGAGATGCGCGCGCGCGCTTGGCGCGCTT 2815
      880 ..... 880
2816 CCCTATTATCCGTTACGCCGCCACTTGGCGAGAAATCCCGTTTCAACAG 2865
      ::|||::: ||| ::|||::: |||
881 .....AsnValThr.....LysTyrAsnThr 887
2866 CTGACGGTAAGCGCAATTAAGGCTCAGGAGACATTCGCTTATATCTC 2915
      |||::: |||::: |||::: |||::: |||:::
888 LeuAsnIleSer...AsnLeuSerGlyAsnGlySerPheTyrTyrLeuTh 903
2916 GGAATCTCTCGGCTACCGCAGCGGCAAAATTAAGCTGCGGCAAACTTCG 2965
      ::|||::: |||::: |||::: |||::: |||:::
903 rAspLeuSerLysAsnGlnGlyAspLysValValAlaThrLysSerAlaL 920

```

```

2966 AAGGCACTTACACCTTGCTGTCAACATAACGGGACAGCAACCCGTAAGT 3015
      ::|||::: ||||| ||| ::::: ::|||::: |||
920 ySgIAsnPheThrLeuGlnValAlaLysAspLysThrGlyIleProThrLys 936
3016 CTGAGCAATTTGACGCTAGTGTGAGGAAAAAGCAACACACCGCTGCCA 3065
      ::|||::: |||||::: ::|||::: |||
937 ...AsnIleuThrLeuPheAspAlaSerAsnAlaThr.....ArgSe 950
3066 AAATCTTAATTTACACCTGCAAAACGAAACACACGTCGATGCCGCGCATGAC 3115
      ::|||::: ||||| ||| ||||| ||||| |||
950 rAsnLeuValValThrLeuAlaAsnGlySerValAspArgGlyAlaTrpL 967
3116 GTTATACGCTTATCCGCAAGACGCGGAGTCCGCTGCATAATCCG... 3162
      ::|||::: |||::: |||::: |||::: |||:::
967 ySTyrlLysLeuArgGluAsnAsnGlyArgTyrAspLeuTyrAsnProGlu 983
3162 ..... 3162
984 ValGluArgArgHisGlnThrValAspThrProSerValGluMetProAs 1000
3162 ..... 3162
1000 nAspMetGlnAlaAspAlaProSerAlaProSerAsnAsnGluIleAla 1017
3162 ..... 3162
1017 LaArgValAspAlaProValProProAlaProAlaProAlaThr 1033
3163 .....GTCAAGACAAAGAGCTTCCGACAAACCTCGGCAAGGC 3200
1034 GlySerAlaMetAlaLysGlnGlnProLysThrArgProAlaGluThrAl 1050
3201 GGGAGAAACA...GAGGCGCCTTGACGGCAAAACAGCAACTTCGCG 3247
      | ||| ||| ::::: ||||| ||||| |||
1050 agLInProThrMetGlnGluThrLysAlaAlaAsnSerThrGluThrAlaP 1067
3248 CCAAAACA.....CAGCGGAAAAAGACACACGCGCAAGCCTT 3285
      |||::: |||||::: ::|||::: |||
1067 rOlySerAspThrThrThrGlnAlaAspThrSerAsnSerGluSerVal 1083
3286 GAGCGG.....CTGAT 3296
1084 ProSerGluThrThrGluThrValAlaGluAsnSerProGlnIleSerAl 1100
3297 TGGGCGCGGCGCATGCCACCGAAAGCGCAAGT..... 3333
      ::|||::: ||| ::|||::: |||
1100 aserValGluLysAsnAlaGluGluAlaThrGluThrProGlnAsna 1117
3334 ....GTTGCCGAACGCGCGCG..... 3351
1117 spGluValAlaLysGluAlaLysProThrValGluThrLysAspGluThr 1133
3352 .....CAGCGAGCGGGGAAATGCCGCGATTAATGACGCGGCA 3389
      |||::: ||||| ||||| ::|||::: |||
1134 AsnGluMetThrGlnSerGlySerGluAsnThrThrGluThrGlnAlaG1 1150
3390 G..... 3390
1150 uAsnLysValSerGlnProThrGluThrAspLysIleAlaThrValGluTr 1167
3391 ..GAAGAGAAAAACGGCTGCAGCGGATTAACACACCGCTTACGCGG 3435
      ||||| ||||| ||||| ::|||::: |||
1167 hrGluGluThrAlaArgValGluLysGlu...GluThrGlnValAlaSer 1182
3436 .....AAACAGCGCGAAGCGGA..... 3453
1183 GlnThrPheProLysGlnGluProGluMetThrLysGlnIleAlaG1 1199
3453 ..... 3453
1199 uProGluThrArgAsnValProIleValAsnAsnLeuGluGluAlaGlnP 1216
3454 ....ACCGCGCGCGCTACC..... 3468

```

```

1216  roglntHrLysProIleHrValAlaThrAlaGluThrProThrSerAsn 1232
3468  .....
1233  SerLysSerAlaGluLysThrGlnProSerGlyGluThrAsnAlaGluPr 1249
3468  .....
1249  ovalThrProValAlaValSerGluAsnGlnProGluAsnThrIleSerG 1266
3468  .....
1266  lnrProThrGluAspThrValValLysValGluThrGluGluThrProLys 1282
3468  .....
1283  ValGluThrGlyGlyThrLysGluAlaProGlnValAlaSerGlnThrSe 1299
3469  .....
1299  rProLysGlnGluGluProGluThrValGlnProGlnAlaGluProAlaG 1316
3497  GCGGGGATTGCCG.....CAACCGCAGCCC 3522
1316  lngLysAsnSerProThrValAsnAsnValGluGluAlaGlnProGlnThr 1332
3523  CAACCG..... 3528
1333  GlnProThrIleValAlaAlaLysGluIleThrAlaProAsnSerAlaG 1349
3528  ..... 3528
1349  nLysGluThrAlaGlnSerValAlaAsnProLysThrAlaGluGlnPro 1366
3528  ..... 3528
1366  alThrValSerThrGluAsnProValValGluAsnProGluAsnThrThr 1382
3529  .....CAACCCCAA..... 3537
1383  GlnProAlaValAsnSerGluAlaValGlnProGluThrAlaThrThrG 1399
3537  ..... 3537
1399  yThrValSerGlnProLysGluAlaSerThrAspGluThrThrValAla 1416
3537  ..... 3537
1416  erThrAspGluThrThrGlyThrSerAlaGluGluThrThrValAlaAsp 1432
3537  ..... 3537
1433  AsnSerGlnAlaSerLysProLysAlaGlySerAlaGlyAspValaSerSe 1449
3537  ..... 3537
1449  rThrProHAsnValAlaGluProAlaValThrGlyGlyGlyAlaGlyAsp 1466
3538  .....CGGACGGCGACCTGATCGCGCTTATGCCATAGC 3573
1466  erAlaValAlaValProLeuAlaGlyAspLeuThrSerThrAsnThrAsn 1482
3574  GGTGTGAGTGAATTTCCGCC.....ACGCTCAACAGCGCT 3608
1483  ValLeuSerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnValAl 1499
3609  TTTCGCCGTCACAGACGAATGGACCGCGTG...TTTGGCGAAGACCGCC 3655
1499  yLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGly 1516
3656  GCNAACCGCTTGTGACAGCGGATCGGGACCAACCAACACACCTTGG 3705
::  |||||::|||::

```

```

1516  lntYrAsnValTrpValSerAsnThrSerMetAsnGluAsnTrpSerSer 1532
3706  CAAGATTCCGGCCTACCGCCACAACCGACCTGCGCAATCGCTAT 3755
1533  AspGlnTrpArgArgPheSerSerLysSerAlaGlnThrGlnLeuGlyTr 1549
3756  GCAGAAAACCTGGCAGCGGG...CGCGTCGGCATCCGTGTTGGACA 3802
1549  pasrGlnThrIleSerAsnAsnValGlnLeuGlyGlyIleLeuThrTrp 1566
3803  ACCGACCGGAAACACCTTCGACAGCGCATCGCAACTCGCAGCGCTT 3852
1566  alArgAsnSerAsnSerPheAspLysAlaSerSerLysAsnThr...Leu 1581
3853  GCCCAGCGTCGCTTTGCGGCAA...TACGCATCGCAGGTTGCACAT 3899
1582  AlaGlnAlaAsnPheTrpSerLysTrpTrpAlaAspAsnHisTrpTrp 1598
3900  CGGCATCAGCGCGCGGCTTTAGTAGGGGACGCTTCAGACGCGCA 3949
1598  uAlaValAspLeuGlyTrpGlyAsnPheGlnSerAsnLeuGlnThrAsn 1615
3950  TCAGAGCAAAATCCGCGCGCGCTGCTCATTCAGCGCATTCAGCGCA 3999
1615  isAsnAlaLysPheAspArgHisThrAlaGlnIleGlyLeuThrAlaG 1631
4000  TACCGCGCAGGTTTCGCGGATTCGCGATCGAACCGCACATCGCGCA 4049
1632  LysAlaPheAsnLeuGlyAsnValAlaValLysProThrValGlyVal 1648
4050  GCGCTATTTCGTCAAAAGCGATTACGATACGAAACGTCATATTCG 4099
1648  gTrpSerTrpLeuSerAsnAlaAspPheAlaLeuAspGlnAspArgIle 1665
4100  CCACCCCGGCGCTTCGATTCACCGCTACCGCGCGCATTAAGCAGAT 4149
1665  yValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSer 1681
4150  TATTCATTCAACCGCGCGCACATTCATCGACCTTATTCAGCT 4199
1682  TrpThrTrpLysAsnLeuGlyGlu...PheAlaIleThrProIleLeuSer 1697
4200  GTCTATACCGATCCGCGCTTCGCAAAAGTCCGACGCGCGCATACG 4249
1697  argTrp...AspAlaAsnGlnGlyAsnGlyLysIleAsnValSerGly 1713
4250  CCGTATTCGCGCAGGATTCGCAAAACCGCGAGTCGGAATGGCGCTA 4299
1713  yAlaPheAlaTrpAsnValGluAsnGlnGlnGlnTrpAlaAlaGlyLeu 1729
4300  AACGCCGAATCAAGGTTTCAGCTGTCCTCCACGCTGCCCGGCCAA 4349
1730  LysLeuLysTrpHisAsnValLysLeuSerLeuMetGlyGlyLeuThr 1746
4350  GGGCGCGCAATTCGAAGCGCACAGCGCGGCGCAATTCAGGCTAC 4398
1746  salLysGlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1762

seq_name: sp_bacteria:030575
seq_documentation_block:
ID 030575 PRELIMINARY; PRT; 997 AA.
AC 030575;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
SEQUENCE FROM N.A.

```

RC STRAIN-24099;
 RX MEDLINE=98010345; PubMed=9350862;
 RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
 RA Achman M.;
 RT "Clonal descent and microevolution of *Neisseria meningitidis* during 30
 RT years of epidemic spread";
 RL Mol. Microbiol. 25:1047-1064(1997).
 DR EMBL, AF012210; AAC45793.2; -;
 DR MEROPS; S06.001; -;
 DR InterPro; IPR002195; Dihydroorotase.
 DR InterPro; IPR000710; IGA_S6.
 DR Pfam; PF02395; IGAL; 1.
 DR PRINTS; PR00921; IGASEPTASE.
 DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 KW Protease.
 FT NON_TER 1 1
 FT 997 997
 SQ SEQUENCE 997 AA; 110152 MW; 4DE40DE594E5E28E CRC64;

alignment_scores:
 Quality: 1398.00 Length: 1194
 Ratio: 2.047 Gaps: 38
 Percent Similarity: 57.203 Percent Identity: 31.240

alignment_block:
 US-09-303-518D-653 x 030575 ..

Align seg 1/1 to: 030575 from: 1 to: 997

```

94  CTTGCTGTCGGCATCTGCTGCCAAGCCGCGGAGACACATTATTTTCGG 143
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1  LeuAlaIeThrAlaLeuThrProTyrSerGluAlaIeAlaLeuValAspAs 17
144 CATCAACTACCACTACTATCGCGACTTTCGCCGAAATTAAGCGAGTTTG 193
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
17  ValAspTyrGlnIlePheArgAspPheAlaGluAsnIleGlyLysPheP 34
194 CAGTCGGCGGCAAGATATTGAGTTTACACAAAAGGAGGAGTTGGTC 243
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
34  heValIeIyAlaThrAspLeuSerValLysAsnIleGlyIleAsnIle 50
244 GGCCAATCGATGACGAAGCCCGCATGATTTTCTGTGTATTCGCG 293
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
51  GlyAsnAlaIeLeuSerAsnValPrometIleAspPheSerValAlaAsp 67
294 TAAC..GGCGTGGCGCATTTGGCGGCATATATATTGACGCGTGG 340
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
67  IAsnATGATgThrLeuThrValIleAspProGlnTyrAlaValSerVal 84
341 CACATTAACGGCGCGC.....TATACAAATGTTGATTTTGGT 375
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
84  yshIshValIySglYAspGluIleSerTyrTyrGlyshIshAsnGlyHis 100
376 GCGGAGGGAAGCAATCCCGATCAGCACCGCTTTTCTACCAATTTGAA 425
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
101  LeuAspValSerAsnAspGluAsn.....GluTyrArgSerValAla 114
426 AAGAATAATTAATAAGCAGGACATACCGCCATCTTATGGCGG.... 471
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
114  agIAsnAspTyrGluProAsnIySAsnIyPheHis..HisGlyAsnGly 130
472 .....GATTATCATATGCCGCTTTCGACAAATTTGTCACAGAT 510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
130  LyrArgLeuGluAspTyrAsnMetAlaArgLeuAsnIySpheValIleArg 146
511 GCAGAACCTGTTGAGATGACCAAGTTATGATGGTGAAA...TACGC 557
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
147  ValAlaProIleAlaProThrSerAlaGlyGlyValGluIleThrTyrI 163
558 TGATTTAAATAATACCTGATGCTGTGCAATCGACGACGAGACAT 607
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
163  sAspLysAsnArgPheSerGluPheValArgValIleGlyIleThrGln 180

```

```

608  ATGGCGCGTGTGATGAAGACCAACCAATACCGGGAAGTTATATCAT 657
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
180  he.....GluTyrAsnSerArgTyrAsnMetIleArg 190
658  ATTGCAGCGCATATTTCTGGCTGCTGCGTGCATATCCTTT..... 699
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
191  LeuSerArgAlaIleArgTyrAlaIleAlaGlyThrProTyrGlnAsp 207
700 .....GCACAAATGATGATCAGCTGGTGCACAGTCACTTGTGACG 742
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
207  IAsnValThrSerAsnLeuAsnGlnGluIleGlyPheGlyAsp 224
743  AAAAAATTAACATAGCCCA.....TATGCT 768
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
224  snSerLysIshIshSerProGluIySLeuIySglValLeuSerGlnAsn 240
769  TTTTACCAACAGAGGCTCATTTGGCGACATGGCTCACCAATGTTTAT 818
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
241  AlaLeuThrAsnTyrAlaValIleGlyAspSerGlySerProLeuPheAl 257
819  CTATGATGCCCAAGCAAAAGTGTATTAATGAGGTTCGCAACAG 868
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
257  aTyrAspLysGlnGluLysArgTyrPheLeuGlyAlaTyrAspTyr 274
869  GCACCCCTATATAGCAAAAGCAATGCTTCCAGCTAGTTCGTAAGAT 918
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
274  rPalaGlyTyr....GlnLysAsnSerTrpGln.....Glu 284
919  TGG...TTCTATGATGAATCTTGTGAGATACCATTCAGTATTTCTA 965
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
285  TrpAsnIleTyrLysLysGluPheAlaAspGlu.....I 296
966  CGAACCAATCAATGAATGGAATCTTTTACGACATATATATAGCG 1015
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
296  eLysGlnHisAspAsn.....A 302
1016  CAGCAAAATCGATGCCAACAATTAACACATATCTCTACTTATAGATTA 1065
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
302  IeGlyThrIleLysGlyTyrArgGlnHis.....TrpAsnIle 315
1066  AAAACAGACACCGCTTCATTTGATGTTCTTTATCCGACGACGACAG 1115
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
316  ThrSerGlyThr.....AsnSerLysIleLysSerTrpAla.. 327
1116  AGAACCTGTTATCATGCTGCGAGGTGGGTCACAGTATGACCCAGAC 1165
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
328  .....ValArgLeuAlaGlyAsnGluAsn.....A 337
1166  TGAATTAATGAGAAATATTTCTTTATTTGACAAAGGAAAGTGAATTG 1215
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
337  IAsnAsnGlyGlnAsnValThrPheGluAsp.....AsnGlyThrLeu 351
1216  ATACTTACACACATCAACCAAGGCGGCGGCTTGTATTTTGAGGG 1265
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
352  ValLeuAspGlnAsnIleAsnGlnIyAlaIyLysLeuPhePheLysG 366
1266  TAATTTACGCTGCTGCTTAATAAACAGAA...ACGTGGCAAGCGCGG 1312
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
368  yAsnTyrThrValIySglYAlaAsnAsnAspIleThrTrpLeuIyAlaG 385
1313  GCGTTCATATCAGTGAAGGACGTACCGTCTTACTTGAAGTAAACGGCG 1362
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
385  IyIleAspValAlaAspLysLysValValTrpGlnValLysAsnPro 401
1363  GCAAGACGCGCGCTGCCAATAATGCGCAAGGCGGCTGCTTACAG 1412
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
402  AsnGlyAspArgLeuAlaIySglYAlaIySglYThrLeuGlnIleAsn 418
1413  CAAAGGGAACCAAGGCTGCTGAGCGTGGCGAGCGTAAAGTCACTCT 1462
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:
418  yThrGlyValAsnGlnGlyGlnLeuLysValGlyAspPolYThrValIle 435

```



```
819 CTATGATGCCCAAAAGAGGTTAATTAATGCGGTATTCGAACAG 868
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
257 aTyAspLysGlnGlnLysAlaTrpAlaPheLeuGlyAlaTyAspTrp 274
869 GCACCCCTATATGCAAAAGCATGCGTCCAGCTAGTTCGTAAAGAT 918
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
274 rPalagLysr.....GlnLysAsnSerTrpGln.....Gln 284
919 TGG...TTCTATGATGAAATCTTTGCTGGAGATACCATTCACTATTC 965
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
285 TrpAsnLeuTrpLysLysGlnPheAlaAspGluIleLysGln..... 298
966 CGAACCATCAATAATGGAAATACCTTTTAAAGACAAATAATAATGGCG 1015
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1016 CAGGAAATAATCGATGCCAAACATAAACATATCTCTACCTTATGATTA 1065
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
302 laeLysThrIleLysGlyTrpGlyGlnHis.....Trp 313
1066 AAACACAGAACCCGTTCATGTTTATGTTTCTTTATCCGAGACAGCAG 1115
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 LysThrThrGlyThr.....AsnSerHisIleGlySerThrAlaVala 327
1116 AGAAGCTGTTTATCATGCTGCAGAGTGGGTCAACAGTTATCGACCCAGAC 1165
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
327 laArgLeuAlaGlyAsnGluArgGlyAla..... 336
1166 TGAATAATGAGAAATATTTCTTATTAATGACAAAGAAAGTGAATG 1215
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 ..AsnAsnGlyGlnAsnValThrPheGluAsp.....AsnGlyThrLeu 350
1216 ATACTTACGACCAATCATCAACCAAGCGCGCGGCTTTGTTTATGAGAG 1265
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 ValLeuAspGlnAsnIleAsnGlnGlyAlaGlyLeuPhePheLysG 367
1266 TATTTTACGGTCTGCGCTAAAAACACAA...ACGTGGCAAGCGCGG 1312
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 yAspTrpThrValLysGlyAlaAsnAspIleThrTrpLeuGlyAlaG 384
1313 GCGTTCATATGATGATGAGTACGCTACGCTTACTTGAAGTAAAGCGCG 1362
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
384 yLleAspAlaAlaAspGlyLysValValThrGlnValLysAsnPro 400
1363 GCACAGACCGCTGTCCAAATCGCAAGGCGACGCTGCTGTTCAAGC 1412
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 AsnGlyAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnG 417
1413 CAAGGCGAAACCAAGGCTGCTGCGTGGCGAGCGTAAAGTCACT 1462
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
417 yThrGlyValAsnGlnGlyGlnLeuLysValGlyLysGlyThrValIle 434
1463 TAGATCAGCAGCGGAGCATCAAGCAAAAAACAAGCCTTATGTAATC 1512
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 euAsnGlnGlnAlaAspAlaAspLysValGlnAlaPheSerGlnVal 450
1513 GCGTGTGACGCGGAGCGGCGTGCAGTGAATGCCGATCACT 1562
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
451 GlyIleValSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnI 467
1563 CAACCCGCAAACTTATTTGCGCTTCCGCGGCGAGCTTGGATTGA 1612
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
467 euAsnProAspAsnLeuTrpPheGlyPheArgGlyLysArgLeuAspAla 484
1613 ACGGCGATTCGCTTCTCCACCGCATTAATAATCCGATGAAGGCGG 1662
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
484 snGlyAsnAspLeuThrPheGlnHisIleArgAsnValAspGluGlyAla 500
1663 ATGATTGTCAACACATCAAGACAAAGATCCACCGTTACCATTTACGG 1712
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 ArgIleValAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrG 517
1713 CAATTAAGATATTACTACA..... 1731
517 yLysSerLeuIleThrAlaProGlnAsnLeuSerValTyArgLuuIleArg 534
1731 ..... 1731
534 snAspTrpAspAspAspTrpTyTrpGlyTyTrpSerTyArgLysPro 550
1731 ..... 1731
551 IleProGlnGlyLysAspLeuTyTrpLysAsnTyArgTyTrpAlaLe 567
1732 .....ACCG 1735
567 uLysSerGlyLysSerValAsnAlaProMetProGlnAsnGlyGlnThrG 584
1736 GCATTAAC..... 1743
584 LuAsnAsnAspTrpValPheMetGlyTyTrpLysGlnGlnGluAlaGlnLys 600
1744 .....ACCACTTGATAGCAAAAAGAAATTCCTTACAAACGTTG 1784
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 AsnAlaMetAsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPh 617
1785 GTTGGCGAAGATGCAACCAACGAGCGGCGCTCAATCTGAT 1834
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
617 ePheGlyGlnGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsn 634
1835 ACCAAGCGGAAGAGCGGATGCGACTTACTGCTTCGCGCGGAACAAT 1884
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
634 heAsnGlyLysSerAlaGlnAsnArgPheLeuThrGlyGlyThrAsn 650
1885 TTAACGGCAATATCAGCGCAACCAACGCAAACTGTTTTCACGCGCAG 1934
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 LeuAsnGlyLysIleSerValThrGlnGlyAsnValLeuLeuSerGlyAr 667
1935 ACCGACACCGCAGCGCTTACATCATTTAGGAAGCGGCTGCAAAATGG 1984
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
667 gProThrProHisAlaArgAspPheValAsnLysSerAlaArgLysA 684
1985 AAGT.....ATCCCAAGAGCAAAATCGTGTGGCAACGATTTGATC 2028
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
684 sPAlaHisPheSerLysAsnAsnGluValAlaPheGluAspAspTrpIle 700
2029 GACCGCATTTAAGCGGAAACTTCATATT...CAGGCGGACAGC 2075
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
701 AsnArgThrPheLysAlaThrGluIleAlaValAsnImSerAlaSerPh 717
2076 GGTGGTTCCCGCAATGTGCCAAAGTGAAGCGGATTTGCATTTAAGCA 2125
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
717 eSerSerGlyArgAsnValSerAsnIleThrAlaAsnIleThrAlaThrA 734
2126 ATCCGCGCAAGCAGTTTGGGTGTGCGACCGCATCAACCCACACATC 2175
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
734 sPAsnAlaLysValAsnLeuGlyTrp.....LysAsnGlyAspGluVal 748
2176 TGTAACAGTTTCGACTGAGCGGTGTGACAAGTTGTACGCAAAAACAT 2225
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
749 CysValAlaTrpSerAspTrpTyThrGlyValThrCysAsnThrAspAsn 765
2226 TACCGAGATTAAGTGATTCTTCATTTGAGCAAGACGACATGAGAGCA 2275
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
765 user...AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyA 781
2276 ATGCACGCTTGGCGATCAGCTCATTTAAATTCACAGAGACTTGGCACA 2325
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 snValAsnLeuSerGlnAsnAlaAlaLeuValLeu..... 792
2326 CTCACGCGCAATCTTAGTGAGCGGAGACACGCACTATACGTTACGG 2375
792 ..... 792
2376 CAACGCGCACCAAAAGCGCACTCAGCTCGTGGCAATGCCCAAGCA 2425
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


376 GCGAGGAGCAATCCCGATCCGATCCGACCGCTTTTCTACCAATGTGAA 425
101 LeuAspValSerAsnAspGluAsn.....GluTyrArgSerValAl 114
426 AAGAAATTAATTAATCAAGGAGGATCAAGGCGATCCTTATGGCGG 471
114 aglnAsnAspTyrGluProAsnLysAsnThrPhe...HisGlyAsnGlnG 130
472GATTATCATATGCCCGCTTTTGCACAAATTTGTACAGAT 510
130 LysArgLeuGluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGlu 146
511 GCAGAACCTGTTGAGATGACCACTTATATGATGGGTGGAA...TACCG 557
147 ValAlaProLeuAlaProThrSerAlaGlyGlyValGluThrTyrIly 163
558 TGAATTAATTAATTAATCAAGGAGGATCAAGGCGATCCTTATGGCGG 607
163 sAspLysAsnArgPheSerGluPheValArgValAlaGlyAlaGlyThrGlnP 180
608 ATGGGGGCTGTGATGAAGACCAATTAACCGCAAGTTCATTCAT 657
180 he.....GluTyrAsnSerArgTyrAsnMetThrGlu 190
658 ATTCGAGCGCATATTTCTGGCTGCTGGTGGCAATACCTT..... 699
191 LeuSerArgAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspIy 207
700GCACAAATGATCAAGTGTGGCAGCAGTCAACTAGTAGGCG 742
207 LAsnValThrSerAsnLeuAsnGlnGlnGlyLeuIleGlyPheGlyAspA 224
743 AAAAAATTAACATAGCCCA.....TATGCT 768
224 snSerLysHisSerProGluLysLeuLysGluValLeuSerGlnAsn 240
769 TTTTTCACCAAGAGGCTCATTTGGCAGCAGTGGCTCAATGTTTAT 818
241 AlaleuThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAl 257
819 CTATGATGCCAAAGCAAAAGTGTATTAATGGGTATTCACAAACG 868
257 aTyrAspLysGlnGlnLysArgTyrAlaPheLeuGlyAlaTyrAspTyr 274
869 GCACCCCTATATAGGAAACCAATGCTCCAGCTAGTGTGAAGAT 918
274 rPalaglyTyr.....GlnLysAsnSerTrpGln.....Glu 284
919 TGG...TTCTATGATGAATCTTTGCTGGAGATACCCATTCAGTATCTTA 965
285 TrpAsnIleTyrLysLysGluPheAla..... 293
966 CGAACCAATCAAAATGGAATACTTTTAAACGACATATATGCGG 1015
294AspLysIleLysGlnAlaArgAspAsn.....A 302
1016 CAGGAAAAATCGATGCCAACAATTAACACTATCTCTACCTATAGATTA 1065
302 laglyThrIleLysGlnLysGlnLysHis..... 312
1066 AAAACGACGACCGTTCATGTTTAAATGTTTCTTATCCGAGACGCAAG 1115
313TrpAsnIleThrPheGlyThrAsnSerLys 322
1116 AGAAGCTGTTTATCATGCTGCAGGTGGGTCAACAGTTTTCACACCGACG 1165
322 s.....IleGlySerThrAlaValArgL 330
1166 TG.....AATTAATGAGAAATATTTCTTAT 1194
330 euAlaGlyAsnGlnLysAspAlaAsnAsnGlyGlnAsnValThrPheGlu 346
1195 GACAAAGGAAAAAGTGAATGTGATCTTACAGCAACATCAACGAGCGC 1244
347 Asp.....AsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyAl 361
1245 GGGCGGTTTGTATTTTGAAGGATTAATTTACGGTCTCGCTCAAAACG 1294
361 aglyLysLeuPhePheLysGlyAspTyrThrValLysGlyLysAsnAsn 378
1295 AA...ACGTGGCAAGCGCGGCGCTTCATATCAGTATGAGGACGATTCG 1341
378 spIleThrThrLeuGlnLysAlaGlyIleAspValThrAspGlyLysVal 394
1342 ACTGGAAGTAACGCGGTGGCAACGAGCGCTGCTCAAAATCGGCA 1391
395 ValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLys 411
1392 AGGACGCTGCTGTTCAAGCCAAAGGGAACCAAGGCTCGGTCAAGC 1441
411 sglyThrLeuGlnLysAsnGlnLysThrGlyValAsnGlnGlyGlnLeuLysV 428
1442 TGGCGAGCGTAAGTCACTTACATCAGCAGCGGAGCATCAAGCAAA 1491
428 alGlyAspGlyThrValIleLeuAsnGlnGlnAlaAspAlaAspLysLys 444
1492 AAACAAAGCCTTATGAAATCGGCTGTGTCAGCGGAGGAGGAGGTCGA 1541
445 ValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThrLeuVa 461
1542 ACTGAATCCGATATATCAGTTCAACCCGACAAACCTATTTTCGCTTTC 1591
461 IleuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyrPheGlyPheA 478
1592 GCGGCGAGCTTTGATTTGAACGGGCGATTCGCTTGTTCACCGCAT 1641
478 rGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGlnHisIle 494
1642 CAAATACCGATGAAGGGCGATGATGTCAACACATCAACGAAAGA 1691
495 ArgAsnValAspGlnGlyAlaArgIleValAsnHisAsnThrSerHisAl 511
1692 ATCCACGCTTACCATTTACAGGCAATTAAGATATTAAT 1728
511 aSerThrIleThrLeuThrGlyLysSerLeuIleThrAsnProAsnSerL 528
1728 1728
528 euSerValHisSerIleGlnAsnAspTyrAspGluAspPyrSerTyr 544
1728 1728
545 TyrTyrArgProArgArgProIleProGlnGlyLysAspLeuTyrTyrLy 561
1728 1728
561 sAsnTyrArgTyrTyrAlaLeuLysSerGlyGlySerValAsnAlaPro 578
1729ACAACGCGCAATAAC..... 1743
578 etProGluAsnGlyValThrGlnLysAsnAsnSprTrpValPheMetGlyTyr 594
1744ACAACCTTGGATAGCAAAA 1763
595 ThrGlnGlnGlnAlaLysLysAsnAlaMetAsnHisLysAsnAsnGlnAla 611
1764 AGAATTTGCTTACAAAGCTTGTGGTGGCAGAAAGATGACCAAAACGA 1813
611 gIleSerGlyPheSerGlyPhePheGlyGlnGlnLysGlnLysAla 628
1814 ACGGCGGCTCAATTCGATTAACCAACCGGAAGGAGCGGATCGACTTGA 1863
628 snGlyAlaLeuAsnLeuAsnPheAsnGlnLysSerAlaGlnAsnArgPhe 644
1864 CTGCTTTCGCGGAGACAAATTTAAACGGGCAATATTCAGGCAAAACG 1913


```

: ||| |||||.....||| ||||| |||||
417 yThGlyValasnGlnGlnLeuLysValGlyAspGlyThValIleL 434
1463 TAGATTCAGAGCGGAGCATCAAGCAAAAGAGCTTGTAGTAATC 1512
|||||.....| ||| |||||.....|
434 euasnGlnLysAlaAspSerasnGlnLysValGlnAlaPheSerGlnVal 450
1513 GGCTTGGTCAGCGGAGGAGCGGTGCACATGATGCCATATCAAGT 1562
|||||.....| ||| |||||.....|
451 GlyIleValSerGlyArgGlyThLeuValLeuasnSerSerasnGlnI 467
1563 CAACCCGACAAACCTATTCGGCTTCGGCGGAGCATTTGGATTGCA 1612
|||||.....| ||| |||||.....|
467 eaSnPrkAsnLeuThyrPheGlyPheArgGlyGlyArgLeuAsnAla 484
1613 ACGGCGCATCGCTTCGTCACCGCATTCAAATACCGATGAAGGCGC 1662
|||||.....| ||| |||||.....|
484 snGlyAsnAspLeuThrPheGlnHisIleArgAsnValAspGlnGlyAla 500
1663 ATGATTCACACCAATCAAGACAAAGAAATCCAGCTTACCATTACAGG 1712
|||||.....| ||| |||||.....|
501 ArgIleValAsnHisAsnThrGlyHisThrSerThrIleThrLeuThrG 517
1713 CAATAAAGATATTACTACA..... 1731
|||||.....| ||| |||||.....|
517 yLysSerLeuIleThrAsnProAsnSerLeuSerValHisSerIleGln 534
1731 ..... 1731
534 snAspTyrAspGlnAspAspTyrSerTyrTyrArgProArgArgPro 550
1731 ..... 1731
551 IleProGlnGlyLysAspLeuTyrTyrLysAsnTyrArgTyrAlaIle 567
1732 .....ACCG 1735
567 uLysSerGlyLysValAsnAlaProMetProGlnLysnGlnIleThrG 584
1736 GCAATATAC..... 1743
|||||
584 IuAsnAsnAspTrpIleLeuMetGlySerThrGlnGlnIuAlaLysLys 600
1744 .....AACCACTTGATAGCAAAAGAAATGGCTTACAAACGGTTG 1784
|||||.....| ||| |||||.....|
601 AsnAlaMetAsnHisLysAsnAsnGlnIleArgIleSerGlyPheSerGlyPh 617
1785 GTTTCGCGAGAAAGATGCAACCAAAAGACAGCGGCTCAATGAT 1834
|||||.....| ||| |||||.....|
617 ePheGlyGlnGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPr 634
1835 ACCAAGCGAGAGAGCGGATGCGCACTTACTGCTTCGCGGAGCAAAAT 1884
|||||.....| ||| |||||.....|
634 heAsnGlyLysSerIleGlnAsnArgPheLeuLeuThrGlyLysIleAsn 650
1885 TTAAGCGCAATATGACGCAAAAGCAAGCAAGCTTTTCAGCGGACG 1934
|||||.....| ||| |||||.....|
651 LeuAsnGlyLysIleSerValThrGlnGlnLysValIleLeuLeuSerGlyAl 667
1935 ACCGACCGCGACGCTTCAATCATTTAGGAAGCGGCTGCAAAAGATGG 1984
|||||.....| ||| |||||.....|
667 gTrotThrProHisAlaArgAspPheValAsnLysSerSerAlaArgLys 684
1985 AAGGT.....ATCCACAGAGAGAAATGCTGTGGACAGCAATGGATC 2028
|||||.....| ||| |||||.....|
684 sPAlaHisPheSerLysAsnAsnGlnValValPheGlnAspAspTrpIle 700
2029 GACCCGCAATTTAAAGCGGAAACTTCCATAT...CAGGCGGAGCAAGC 2075
|||||.....| ||| |||||.....|
701 AsnArgThrPheLysAlaThrGlnIleAlaValAsnGlnSerAlaSerPh 717
2076 GGTGGTTCCCGCAATGTTGCCAAAGTGGAGGCGCATGGCATTTAAACA 2125
|||||.....| ||| |||||.....|
717 eSerSerGlyArgAsnValSerAspIleThrAlaAsnIleThrAlaThr 734
2126 ATACGCCCCAGACAGTTTTCGTTGCGACCGCATCAAAAGCCACACATC 2175
|||||.....| ||| |||||.....|
734 sPAsnAlaLysValAsnLeuGlyTyr.....LysAsnGlyAspGlnVal 748
2176 TGTACAGCTTCGAGCTGACAGCGGTGCACAAAGTTTACGAAAAACCAT 2225
|||||.....| ||| |||||.....|
749 CysValArgSerAspTrpThrGlyTyrValIleThrCysAsnThrGlyAsnLe 765
2226 TACCGACGATAAGTGATGCTTCATTCAGCAAGACCGACATCAAGGCA 2275
|||||.....| ||| |||||.....|
765 user...AspLysAlaLeuAsnSerPheGlyValIleThrGlnIleAsnGly 781
2276 ATGTACGCTTCGCGATTCAGCGCTCAATTAATCTCAGACAGCTTCGACA 2325
|||||.....| ||| |||||.....|
781 snAlaAsnLeuSerGlnAsnAlaIleAlaLeuThrLeu...GlyLysAlaAla 796
2326 CTCACAGCGCAATCTTAGTCCAGGCGGAGACAGCACTATACGGTTACGCG 2375
|||||.....| ||| |||||.....|
797 LeuTrpGlyGlnIleGlnGly..... 803
2376 CAACGCCACCAAAAGCGCAACCTCAGCTCGTGGCAATGCCCAAGCA 2425
|||||.....| ||| |||||.....|
804 .....GlnGlyAsnSerArgValS 810
2426 CATTTATCAAGCCACATTTAAAGCGCAACATCGGCTTCGGACAATGCT 2475
|||||.....| ||| |||||.....|
810 eTLeuAsnGln..... 813
2476 TCATTTAATCTTAAGCAACAAGCGCGTACAAACGGCAGTCAAGCTTTC 2525
813 ..... 813
2526 CGACAAGGCTAAGCAAGGTAAGCCATTCGCGACTCAAGGCAATGTCT 2575
813 ..... 813
2576 CCCTACCGCATTAAGGCAATTCATTTGAAAACAGCGCTTTACCGGA 2625
813 ..... 813
2626 AAATACAGCGCGCGCAAGGATACGGCATTAACCTTAAAGACAGCGAATG 2675
|||||.....| ||| |||||.....|
814 .....HiserLysTr 817
2676 GACGTCGCGTCGCGGACGAGATTTAGCAATTTAACTTGAACAGCGCA 2725
|||||.....| ||| |||||.....|
817 PHisLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerH 834
2726 CCAATTACATCAATTCGCGCTATGACAGAGATGCGGACGCGGCAAGC 2775
|||||.....| ||| |||||.....|
834 IsIleHisLeuAsnAsnAla.....SerAspAlaIleSerAla..... 846
2776 GCGAGTGGCGCAGATGCGCGCGCGCGCTTCGCGCGCTTCCCTAATTATC 2825
846 ..... 846
2826 CGTTACGCGCGCAACTTGGCAGCAATCCGTTTCAACAGCTGACGATA 2875
|||||.....| ||| |||||.....|
847 .....AsnLysTyrHisThrLeuLysIle 855
2876 ACGGCAAAATTTGAACGCTCAGGAGAAATTCGCTTATGTGCGGAACCTTC 2925
|||||.....| ||| |||||.....|
855 sn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuAla 870
2926 GCGTACCGGACGCGCAATTTGAAGCTGCGCGCAAGTTCGGAAGGCACTTA 2975
|||||.....| ||| |||||.....|
871 LysAsnLeuGlyAspLysValIleValLysGlnSerAlaSerGlyLysTyr 887
2976 CACCTTGGCTGTCAACATACCGGCAAGCAAGCCGTAAGTCTTCGACAAAT 3025
|||||.....| ||| |||||.....|
887 rgIleuHisValGlnAspLysThrGlyLysPro...AsnGlnGlnGlyL 903

```

```

3026 TGAACGCTAGTGAAGAAAGACACACACCGCTGTCGAAATCTTAAT 3075
||||| : : : : : ||| : : : : :
903 euasnleupheaspalaserServalArgAspArgSer...HisLeuSer 918
3076 TTACACCTCGAAACGACACATCGATGCCGCGCATGGCTTATCAGCT 3125
: : : : : ||| : : : : : ||| : : : : :
919 ValSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrThrI 935
3126 TATCCGCAAGACGAGCGAGTTCGCGCTGATATCGGTCAAGAACG 3175
: : : : : ||| : : : : : ||| : : : : :
935 eLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnA 952
952 rg.....ArgArgValLysProAlaProSerProAlaThrAsnThr 965
3226 GCAAACGACGACACTTCCGCCCAACACAGCGGAAAGAACAC 3273
||||| : : : : : ||| : : : : :
966 AlaSerGlnAlaGlnThrAspSerAlaGlnIleAlaLysProGlnAsn 981
seq_name: sp_bacteria:Q956X3

seq_documentation_block:
ID Q956X3 PRELIMINARY; PRT; 993 AA.
AC Q956X3;
DC 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OX Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24400.
RA MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
RA Achtman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread.";
RL MOL. Microbiol. 25:1047-1064(1997).
DR EMBL; AF012208; AAC45791.2; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGA1.1
DR PRINTS; PR00821; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 993 993
SQ SEQUENCE 993 AA; 109441 MW; 109FA2EF88AC3C6 CRC64;

```

alignment_scores:

Quality: 1370.00 Length: 1173
 Ratio: 2.066 Gaps: 34
 Percent Similarity: 56.522 Percent Identity: 30.861

alignment_block:

US-09-303-518d-653 x Q956X3

Align seg 1/1 to: Q956X3 from: 1 to: 993

```

94 CTGTGCTGGGATCTGCGCCAGCGGGGAGACACACTTATTGGG 143
||||| : : : : : ||| : : : : :
1 LeuAlaTyrAlaLeuThrProTyrSerGlnAlaIleValArgAspAs 17
144 CATCAACTCAATATCATCGGACTTGGCCGAATAAAGCAGATTGG 193
: : : : : ||| : : : : : ||| : : : : :
17 pValaspyrTyrGlnIlePheArgAspPheAlaGluAsnGlyLysPhe 34

```

```

194 CAGTCGGCGCCAAAGATATATGAGCTTTACACAAAAAGGGAGTTGTC 243
||||| : : : : : ||| : : : : :
34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50
244 GCGAAATGATGACGAAAGCCCGCATGATGATTTTCTGTGTATGCGC 293
||||| : : : : : ||| : : : : :
51 GlnAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67
294 TAAC...GGCTGGCGGATGGCGGCGGATGATATATATGTCAGCGTCG 340
||||| : : : : : ||| : : : : :
67 IAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValI 84
341 CACATAACGGCGGC.....PATAACAATGTTGATTTTGGT 375
||||| : : : : : ||| : : : : :
84 yShIsValLysGlyAspGluIleSerTyrTyrGlnHisAsnGlnHis 100
376 GCGGAGGAGCAATCCGATCAGCACCGCTTTCTTCCAAATTTGTAA 425
: : : : : ||| : : : : : ||| : : : : :
101 LeuAspValSerAsnAspGluAsn.....GluTyrArgSerValAl 114
426 AAGAAATATATTAAGCAGGACTAACGGCATCTTATGCGGC... 471
: : : : : ||| : : : : : ||| : : : : :
114 aGlnAsnAspTyrGluProAsnLysAsnTyrHis...HisGlyAsnGln 130
472 .....GATTATCATATGCCGCGTTGCGACAAATTTGTCAGAT 510
||||| : : : : : ||| : : : : :
130 LysArgLeuGluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGlu 146
511 GCAGAACCTGTTGAGATGACGACGATTATATGATGATGAGAA...TACGC 557
: : : : : ||| : : : : : ||| : : : : :
147 ValAlaProIleAlaProThrSerAlaGlyLysGlyValGluThrTyrLy 163
558 TGATTTAATATATACCTGATCGTTCGATATGGAGCAGCAGACAT 607
||||| : : : : : ||| : : : : :
163 sAspLysAsnArgPheSerGluPheValArgValGlyAlaGlyThrGlnP 180
608 ATTGCGGCTGTGATGACGACGACCAATACCCGAAAGTTCATATCAT 657
||||| : : : : : ||| : : : : :
180 he.....GluTyrAsnSerAlaGlyTyrAsnMetThrGlu 190
658 ATTGCACCGCATATCTTGGCTGCGTGGGCAATACCTTT..... 699
: : : : : ||| : : : : : ||| : : : : :
191 LeuSerArgAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspVa 207
700 .....GCACAAATGATGATGATGCTGGCAGCAGTCACTTACGTACG 742
: : : : : ||| : : : : : ||| : : : : :
207 IAsnValThrSerAsnLeuAsnGlnGlnGlyLeuIleGlyPheGlyAspA 224
743 AAAAAATTAAACATAGCCCA.....TATGGT 768
: : : : : ||| : : : : : ||| : : : : :
224 snSerLysHisHisSerProGlnLysLeuLysGluValLeuSerGlnAsn 240
769 TTTTACCACAGAGGCTCATTTGGCGACAGTGCACCAATGTTTAT 818
||||| : : : : : ||| : : : : : ||| : : : : :
241 AlaLeuThrAsnTyrAlaValIleGlnIleAspSerGlySerProLeuPheAl 257
819 CTATGATGCCCAAAAGCAAAAGTGTATATATAGGGGTATTGCAACAG 868
||||| : : : : : ||| : : : : : ||| : : : : :
257 atYrAspLysGlnGlnLysArgTyrValPheLeuGlyAlaTyrAspTyrT 274
869 GCAACCCCTATATAGAAAGCAATAGCTCCAGTAGTTGTAAGAT 918
||||| : : : : : ||| : : : : : ||| : : : : :
274 rPalaglyTyr.....GlnLysAsnSerIrrpGln.....Glu 284
919 TGG...TTCTATGATGAATCTTGTGAGATACCATTCAGTATTCGA 965
||||| : : : : : ||| : : : : : ||| : : : : :
285 TrpAsnIleTyrLysLysGluPheAla..... 293
966 CGAACCAATCAAAATGGAATAATCTTTTAAAGCAATATATATAGGCG 1015
: : : : : ||| : : : : : ||| : : : : :
294 .....AspLysIleLysGlnArgAspAsn.....A 302

```



```

814 ..... 814
2605 GAAACAGCCGCTTACCGGAAATACAGCGCGGACAGATACCGCAT 2654
814 ..... 814
2655 AACTTAAAGACGAGATGAGCGCTCCGCGGACGAGATTAAGCA 2704
815 ..... HisSerLysThrPheLysLeuThrGlyAspSerLysValHisA 828
2705 ATTTAACTTACAGACCGCCACATTAACATTCAGCTTACGAGT 2754
828 snLeuSerLeuAlaAspSerHisLysLeuHisValHisA.....Ser 842
2755 GATGGCGGACGCGGCAACCGGAGTCCGCGAGATGGCGCGCGCG 2804
843 AsplaglinserAla..... 847
2805 TTCGGCGCGCTCCCTATTATCCGTTACGCGCGCACTTGGCAGATCC 2854
848 ..... Asn 849
2855 GTTTCACACGCTGACGGTAAACGCAATTGAAACGCTACGAGACTTC 2904
849 IntYrHisThrLeuLysLysn...HisLeuSerGlyAsnGlyHisPhe 864
2905 CGCTTATGTCGAACTCTTCGCTACCGGACGCGCAATTAAGCTGGC 2954
865 HisTyrLeuThrHisLeuAlaGluAsnLeuGlyAspLysValLeuVal 881
2955 GGAAGTTCGGAAGGCACTTACACTTGGCTGTACACATTAAGCGCAAG 3004
881 sglSerAlaSerGlyHisTyrGlnLeuHisValGlnAspLysThrGly 898
3005 AACCGCTAGCTCGACCAATTCAGCTAGTGGAGGAAAGCAACACA 3054
898 lupo...AsnGlnGlyLeuAsnLeuPheAspAlaSerValGln 913
3055 CCGCTGCCGAAATCTTAATTCACCTTCGCAAAACGACACGCTGATGC 3104
914 AspArgSer...ArgLeuSerValSerLeuAlaAsnHisValAspLe 929
3105 CGGCGCTAGCTGATTCAGCTTATCCGCAAGACGCGAGTCCGCTGC 3154
929 uGlyAlaLeuArgTyrThrLysThrGluAsnGlyLysThrArgLeu 946
3155 ATATTCGCTCAAGAACAGACGCTTCCGCAAACTCGGACGCGGGA 3204
946 yrasnProTyrAlaGluAsnArg.....ArgArgValLysProVal 959
3205 GAAACAGAGCCGCTTACGCGCAAAACGAGCAACTTGGCGCAACGA 3254
960 ProSerProLysThrAsnThrAlaSerGlnAlaGlnThrAspSerAla 976
3255 ACAGCGGGAAGAAAGCAAC 3273
976 nleAlaLysProGlnAsn 982
seq_name: sp_bacteria:O30574
seq_documentation_block:
ID O30574 PRELIMINARY; PRJ: 996 AA.
AC O30574;
DT 01-JAN-1998 (Tremblrel. 05. Created)
DT 01-JAN-1998 (Tremblrel. 05. Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19. Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=24024;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012206; AAC45789.2; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_s6.
DR Pfam: PF02395; IGA1.1
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 996 996
SQ SEQUENCE 996 AA; 109688 MW; B3ABDFE54C37D9 CRC64;

```

alignment_scores:

```

Quality: 1369.50 Length: 1194
Ratio: 2.023 Gaps: 38
Percent Similarity: 56.700 Percent Identity: 31.072

```

alignment_block:

```

US-09-303-518D-653 x O30574 ..

```

```

Align seg 1/1 to: O30574 from: 1 to: 996

```

```

94 CTGTCGTCGGGCAATTCGCGCAAGCCGCGGGGACACACTTATTGGG 143
|||||..... |||.....
1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17
144 CATCAACTCACTAATCTATCGCACTTGGCCGAATTAAGCAAGTTTG 193
|||||..... |||.....
17 rValAspTyrGlnLysPheArgAspPheAlaGluAsnLysGlyLysPhe 34
194 CAGTCGGCGGCGAAGATATGAGTTTACACAAAAAGGAGTGGTC 243
|||||..... |||.....
34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyLysnLys 50
244 GGCATATCGATGACCAAGCCCGATGATGATTTCTTGGTATCGCG 293
|||||..... |||.....
51 GlyAlaAlaLeuSerAsnValProMetLysPheSerValAlaAspVa 67
294 TTAAC...GGCTGGCGGCAATGGCGGCGATCATATATTTGAGCGTGG 340
|||||..... |||.....
67 lAsnArgAlaGlnThrLeuThrValLysAspProGlnTyrAlaValSerVal 84
341 CACATTAAGCGCGC.....TATCAATATGATTTTGGT 375
|||||..... |||.....
84 yHisValLysGlyAspLeuLysSerTyrTyrGlyHisHisAsnGlyHis 100
376 GCGGAGGAGAACCAATCCGATCAGACCGCTTTCTTACCAATTTGAA 425
|||||..... |||.....
101 LeuAspValSerAsnAspGluAsn.....GluTyrArgSerValAla 114
426 AAGAAATATATTAAGACGAGGACTAAGCGCCATCTTATGGCGC.... 471
|||||..... |||.....
114 aGlnAsnAspTyrGluProAsnLysAsnThrPheHis...HisGlyAsnGln 130
472 .....GATTATCATATGCGCGCTTGGCAATTTGTCACGAT 510
|||||..... |||.....
130 LysArgLeuGluLysPyrLysMetAlaArgLeuAsnLysPheValThrGlu 146
511 GCAGAACCTGTTGAGATGACCATGTTATATGATGGTGGAAA...TACGC 557
|||||..... |||.....
147 ValAlaProLysAlaProLysSerAlaGlyGlyLysValGluThrTyrLys 163
558 TGATTTAATTAATATACCTGATCGTTGCAATGGGACGACGACAAAT 607
|||||..... |||.....
163 sAspLysAsnArgPheSerGluPheValArgValGlyAlaGlyThrGlnP 180

```



```

734 spaenAlaLysValAsnLeuGlyTyr.....LysAsnGlyAspGluVal 748
734 TGTACAGTGTGGACGTGGACGGGTCTACAAAGTTGTACCCAAAAACCAT 2225
749 CysValAlaIleSerAspTyrThrGlyTyrValThrCysAsnThrAspAsnLe 765
2226 TACCGCAATAAAGTGAATGCTTCATTCATGACCAAGACCGACATCAGAGGCA 2275
765 user...AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyA 781
2276 ATGCAAGCTTCGCGATCAGCTCATTTAAATCTCAGAGACTTGGCACA 2325
781 snValAsnLeuSerGlnAsnAlaAlaLeuValLeu...GlyLysAlaAla 796
2326 CTCACGCGCAATCTTAGTCGAGGGGAGACACGACCATATCGATTACGGC 2375
797 LeuThrPglYglnIleGlnGly..... 803
2376 CAACGCCCAACCAACGGCAACCTCAGCCTCTGGGCAATGCCAGCA 2425
804 .....GlnGlyAsnSerArgValS 810
2426 CATTTAATCAAGCCACATTTAAGCGACACATCGGCTTCGACAAATGCT 2475
810 erLeuAsnGln..... 813
2476 TCATTTAATCTAAGCAACAGCCGTACAAAAGCGAGTCTGACGCTTC 2525
813 ..... 813
2526 CGCAACGCTAAGGCAAGCAAGCATTCGCGACTCAACGGCAATGCT 2575
813 ..... 813
2576 CCCTAGCGATAAGCAGTATTCCATTTGAAAACGCGCTTTACCGCA 2625
813 ..... 813
2626 AAAATCAGCGGGCGCAAGATACGGCATTAACCTTAAAAGACAGCGAATG 2675
814 .....HisSerIleStr 817
2676 GACGCTGCGGTCGGGACGAGATTAAGCAATTTAAACCTTGACAAAGCGCA 2725
817 PheIleuThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerH 834
2726 CCATTACACTCAATTCGCGCTATCGACAGATGCGGCGAGCGCGCAAAAC 2775
834 IeIleHisLeuAsnAsnAla.....SerAspAlaGlnSerAla..... 846
2776 GGCAGTGGCGGCAAGATGCGCGCGCGCGCTTCGCCCTTCCCTATTAATC 2825
846 ..... 846
2826 CGTTACGCGCGCAACTGCGCGAGATTCGCTTCAACAGCGCTGACAGGTA 2875
847 .....AsnLysTyrHisThrLeuIleA 855
2876 ACGGCAAAATGAACGTCAGGAGACATTCGCTTATGTCGGAATCTTC 2925
855 sn...HisLeuSerGlyAsnGlnHisPheHisTyrLeuThrHisLeuAla 870
2926 GCGTACCGCGGCAAAATTAAGCTGGCGGCAAGTTCGGAAGCAGCTTA 2975
871 LysAsnLeuGlyAspLysValLeuValLysGlnSerAlaSerGlyHisTyr 887
2976 CACTTGGCTGTCAACAATTAACGCGCAAGAACCGCTAGTGTCCAGCAAT 3025
887 rGlnLeuHisValGlnAspLysThrGlyLupro...AsnGlnGluGlyL 903
3026 TGACGGTAGTGGAAGAAAAGCAACACACGCGCTGTCCGAAATCTTAAT 3075
|||||..... ||| .....

```

```

903 euAsnLeuPheAspAlaSerSerValGlnAspArgSer...ArgLeuSer 918
3076 TTCACCCCTCAAAACGACACACGTCGATGCCGCGCATTCATCAGCT 3125
919 ValSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrThrI 935
3126 TATCCGCAAAAGACGGCGAGTTCGCGCTGCATATTCGCTCAAGAACAG 3175
935 euYsThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnA 952
3176 ACCTTCCGCAAACTCGCGCAGGGGGGAGAAACGAGCGCCCTTGACG 3225
952 rg.....ArgArgValLysProAlaProSerProAlaThrAsnThr 965
3226 GCAAAACAGCGCAAACTTCGCGCCCAACAGACGCGGCAAAAAAGCAACGC 3275
966 AlaSerGlnAlaGlnLysAla.....ThrGlnThrAspGlyAl 978
3276 GCAAAAGCTTGACGGCGTGAATTCGCGCGCGCGCATGCGCAACGAAAGG 3325
978 acIn.....IleAlaLysProGlnAsnIleVal..... 987
3326 CAGAAAGTGTGCGCAACCGCGCGCGCGAGCA 3357
988 .....ValAlaIleProSerProGlnAla 995

```

seq_name: sp_bacteria:Q9S6X4

seq_documentation_block:

```

ID Q9S6X4 PRELIMINARY; PRT; 992 AA.
AC Q9S6X4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ICAI PROTEASE (FRAGMENT).
GN ICAI.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23906;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread.";
RL MBL Microbiol. 25:1047-1064(1997).
DR EMBL; AF012204; AAC45787.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroproctase.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02385; ICAI; 1.
DR PRINTS; PRO0921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROPROCTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 992
SQ SEQUENCE 992 AA; 109239 MW; 9ED0A41019917CB3 CRC64;

```

alignment_scores:

```

Quality: 1364.50 Length: 1166
Ratio: 2.055 Gaps: 34
Percent Similarity: 56.947 Percent Identity: 30.961

```

alignment_block:

US-09-303-518D-653 x Q9S6X4 ..

Align seg 1/1 to: Q9S6X4 from: 1 to: 992

```

94 CTGTCGTTGGCATTTGCCCCAAGCCGGCGGAGACACATTATTTCGG 143
|||||..... ||| .....

```

[illegible]

```

1732 ..... ACCG 1735
567 uLyserGlyGlySerValAsnAlaPrometProGluAsnGlyInThrG 584
1736 GCATTAAC..... 1743
584 luAsnAsnAspTrpIleLeuMetGlySerThrGlnGluAlaLysLys 600
1744 ..... AACACTGGATAGCAAAAAAGAAATTCGCTAACACGGTTG 1784
601 AsnAlaMetAsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPh 617
1785 GTTTGGCGAGAAAGATGCACAAACAAACGAGCGGCTCAATCTGAATT 1834
617 ePheGlyGluGluAsnGlnGlySerGlyHisAsnGlyAlaLeuAsnLeuAsn 634
1835 ACCAAGCGGAAGAGCGGATCGCCTTACTGCTTCCGGCGGACAAAT 1884
634 heAsnGlyLysSerAlaGlnAsnArgPheLeuLeuThrGlyLysAsn 650
1885 TTAAAGCGCAATATCACGCAAAACAAACGCAACTGTTTTCAGCGGAC 1934
651 leuAsnGlyLysIleSerValThrGlnGlyAsnValLeuLeuSerGlyAr 667
1935 ACCGACACCGCGACGCTACATATCTTAGAAGCGGGTGTCAAAATGG 1984
667 gProThrProHisAlaArgAspPheValAsnLysSerSerAlaGlnLysA 684
1985 AAGGT.....ATCCCAAGGAGAAATCGTGGGACACAGATGTGATC 2028
684 sPalaHisPheSerLysAsnAsnGluValAlaPheGluAspAspTrpIle 700
2029 GACCGCACATTTAAAGCGGAAATTCATCATTT...CAGGCGGACAAAG 2075
701 AsnArgThrPheLysAlaThrGluIleThrValAsnGlnSerLaseRph 717
2076 GGAGGTTCGCCGCAATCTGCCAAAGTGGAGCGGATGGCATTTAAGCA 2125
717 eSerSerGlyArgAsnValSerAsnIleThrAlaAsnIleThrAlaThrA 734
2126 ATCAGCGCCCAAGCAGTTTTCGTGCGCACCCGATCAAGGCCACAAATC 2175
734 sPAsnAlaLysValAsnLeuGlyTyr.....LysAsnGlyAspGluVal 748
2176 TGTACACGTTTCGAGTGGAGCGGTGTGACAAAGTTTACGAAAAACCAT 2225
749 CysValAlaSerAspTyrThrGlyTyrValThrCysAsnThrChLysnLe 765
2226 TACCGACGATTAAGTGATGCTTCATTTAGCAAGACCGGACATCAGGCA 2275
765 uSer...AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyA 781
2276 ATGTGACGCTTGGCGATCAGCTCATTTAAATCTACAGGACTTGGCCACA 2325
781 snValAsnLeuAsnGlnAsnAlaAlaLeuValLeu...GlyLysAlaAla 796
2326 CTCAAGCGCAATCTTAGTGCAGGCGGAGACACGACTATACGTTACGGC 2375
797 LeuTrpGlyGlnIleGlnGly..... 803
2376 CAAGCGCACCAAAAGCGCAACCTCAGCTCGTGGGCATGCCCAAGCAA 2425
804 .....GlnGlyAsnSerArgValAs 810
2426 CATTTAATCAAGCCATTTAAAGGCAACACATCGGCTTGGACAAATGCT 2475
810 ePheLeuAsnGln..... 813
2476 TCATTTAATCTAAGCACACAGCCCGTACAAAAGCGCAGTGTGACGCTTTC 2525
813 ..... 813
2526 CGACAAGCTTAAGCAAAAGTAAAGCATTCGCACTCAAGGCATGTCT 2575

```

```

813 ..... 813
2576 CCTACGCGATAAGCGAGTATTCATTTGAAAAACAGCCGTTTACCGGA 2625
813 ..... 813
2626 AAAATAGCGGCGGACAGATACGGCATTAACCTTAAAAAGACAGGAAAT 2675
814 .....HisSerLysTr 817
2676 GAGCGTCCGCGGCGGCAATTAAGCAATTTAAACCTTGACAAAGCCA 2725
817 pHisLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerH 834
2726 CCATTAAGCTCAATTCGCGCTATGCACACGATGCGGACGCGCGCAACC 2775
834 IsIleHisLeuAsnAsnAla.....SerAspAlaGlnSerAla..... 846
2776 GGCAGTGGCGGACAGATGCGCGCGCGCGCTTCGCGCCGTTCCCTATTATC 2825
846 ..... 846
2826 CGTTACGCGCCCACTTCGCGAGAAATCCGTTTCAACAGCGTAGA 2875
847 .....AsnGlnIleHisThrLeuLysIleH 855
2876 AGCGCAATTTGAAGGTCAAGGAGACATTCGCTTATGTGCAACTCTTC 2925
855 sn...HisLeuSerGlyAsnGlnHisPheHisTyrLeuThrHisLeuAla 870
2926 GCGTACCGGCGCAAAATTAAGCTGGCGGAAATTTCCGAAGGACSTTA 2975
871 GluAsnLeuGlyAspLysValLeuValLysGluSerAlaSerGlyHisTyr 887
2976 CACTTGGCTGTCAACAATACCGGCAACACCCGTAAGTGCATGCACT 3025
887 rGlnLeuHisValGlnAspLysThrGlyGluPro...AsnGlnGluGlyL 903
3026 TGACGCTACTGGAAGAAACAAACACACACCGCTCTCCGAAATCTTAAT 3075
903 euAsnLeuPheAspAlaSerSerValAlaArgAspArgSer...HisLeuSer 918
3076 TTCACCTCGCAAAAGCAGTGCAGTGCAGCGGCGGATGCGTTTCAGCT 3125
919 ValSerLeuAlaAsnHisHisValAspLeuGlyAlaLeuArgTyrThrI 935
3126 TATCGCAAGACGCGGAGTTCCGCTGCATTAATCCGTTCAAAAGACAAAG 3175
935 eLysThrGluAsnGlyIleThrArgLeuTyrAsnProThrAlaGluAsnA 952
3176 AGCTTTCGCAAAACTCGCGCAAGCGGAGAAACAGAGCGCGCTTGACG 3225
952 rG.....ArgArgValLysProAlaProSerProAlaThrAsnThr 965
3226 GCAAAACAGCACAACTTGGCGGCAACACAGCGGAGGAAAGAACAGCANC 3273
966 AlaSerGlnAlaGlnThrAspSerAlaGlnIleAlaLysProGlnAsn 981

```

seq_name: sp_bacteria:Q956X5

seq_documentation_block:

ID Q956X5 PRELIMINARY; PRT; 992 AA.

AC Q956X5;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE ICA1 PROTEASE (FRAGMENT).

GN ICA.

OS Neisseria meningitidis.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=4877;

RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491.
 RA MEDLINE=98010345; PubMed=9350862;
 RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
 RA Achman M.;
 RT "Clonal descent and microevolution of *Neisseria meningitidis* during 30
 RT years of epidemic spread."
 RT Mol. Microbiol. 25:1047-1064(1997).
 RL EMBL AF012203; AAC45786.1; -.
 DR MEROPS; S06.001; -.
 DR InterPro; IPR002195; Dihydroorotase.
 DR InterPro; IPR000710; IGA_S6.
 DR Pfam; PF02395; IGA1; 1.
 DR PRINTS; PR00921; IGASRPASE.
 DR PROSITE; PS00462; DIHYDROOROTASE_1; UNKNOWN_1.
 DR PROTEASE.
 FT NON_TER 1 1
 FT NON_TER 992 992
 SHORCE 992 AA; 109239 MW; 306FA51B7213D49 CRC64;

alignment_scores:
 Quality: 1363.50 Length: 1166
 Ratio: 2.057 Gaps: 34
 Percent Similarity: 56.861 Percent Identity: 30.961

alignment_block:
 US-09-303-518D-653 x Q956X5 ..

Align seg 1/1 to: Q956X5 from: 1 to: 992

```

94 CTGTGTCGCGCATCTGCCCCAGCCGCGGAGACACTTATTTTCG 143
|||||:|||||: ||| :|||:|||||: |||:
1 LeuAlaIyRAlaLeuThrProTyrSerGluAlaIaLeuValArgAspAs 17
|||||:|||||: |||:|||||: |||:|||||: |||:
144 CATCAACTACCAATCTATCGCGACTTTGCCGAAATTAAGCAAGTTTG 193
|||||:|||||: |||:|||||: |||:|||||: |||:
17 pValAspTyrGlnIlePheArgSpPheAlaGluAsnLysGlySphP 34
|||||:|||||: |||:|||||: |||:|||||: |||:
194 CAGTGGGGCGGAAAGATATTTAGCTTACCAAAAGCGAGTGTGTC 243
|||||:|||||: |||:|||||: |||:|||||: |||:
34 heValGlyValAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50
|||||:|||||: |||:|||||: |||:|||||: |||:
244 GGCNAATCGATGACGAAAGCCCGGATGATGATTTTCTGTGTCGTCGCG 293
|||||:|||||: |||:|||||: |||:|||||: |||:
51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67
|||||:|||||: |||:|||||: |||:|||||: |||:
294 TTAAC..GGCGTGGCGCATTTGGCGGCGATCATATATTTGACCGCTGG 340
|||||:|||||: |||:|||||: |||:|||||: |||:
67 IAsnArgArgThrLeuThrValIleAspProGlnIleValAlaValSerVal 84
|||||:|||||: |||:|||||: |||:|||||: |||:
341 CACATTAACGGCGGC.....TATTAACATGTTGATTTTGGT 375
|||||:|||||: |||:|||||: |||:|||||: |||:
84 yHHisValLysGlyAspGluIleSerTyrTyrClyHHisAsnGlnLys 100
|||||:|||||: |||:|||||: |||:|||||: |||:
376 GCGGAGGAGCAATCCGATCAGACCGCTTTCTTACCAATTTGTGA 425
|||||:|||||: |||:|||||: |||:|||||: |||:
101 LeuAspValSerAsnAspGluAsn.....GluTyrArgSerValAl 114
|||||:|||||: |||:|||||: |||:|||||: |||:
426 AAGAATATTTATTAAGCAGACGACTTAAGCGCATTCCTTATGGCGGC 471
|||||:|||||: |||:|||||: |||:|||||: |||:
114 aGlnAsnAspTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnG 130
|||||:|||||: |||:|||||: |||:|||||: |||:
472 .....GATTATCATATGCCCGCTTTCACAAATTTGTCAAGAT 510
|||||:|||||: |||:|||||: |||:|||||: |||:
130 IArgLeuGlnAspTyrAsnMetAlaArgLeuAsnLysPheValThrGlu 146
|||||:|||||: |||:|||||: |||:|||||: |||:
511 GCAGAACTCTGTGAGTACACAGTATATGATGGGTGAAA...TACGC 557
|||||:|||||: |||:|||||: |||:|||||: |||:
147 ValAlaProIleAlaProThrSerAlaGlyGlyValGluThrTyrI 163
|||||:|||||: |||:|||||: |||:|||||: |||:
558 TGATTTAAATTAATCCCTGATCGTGTTCGATCGGAGCAGACGAT 607
|||||:|||||: |||:|||||: |||:|||||: |||:

```

```

163 sasPLysAsnArgPheSerGluPheValArgValGlyAlaGlyThrGlnP 180
608 ATTTGGCGGTCTGATGAGACGACCAATTAACCGGAAATGATCATCAT 657
|||:|||||: |||:|||||: |||:|||||: |||:
180 he.....GluTyrAsnSerArgTyrAsnMetThrGlu 190
658 ATTTGACGCCATATTTGGCTCGTGGTGCATTAACCTTT..... 699
|||||:|||||: |||:|||||: |||:|||||: |||:
191 LeuSerArgAlaIyRArgTyrAlaIleAlaGlyThrProTyrGlnAspVa 207
700 .....GCACAAATATGATCAGGCTGGCGACAGTCACTTAAGTAAAGG 742
|||||:|||||: |||:|||||: |||:|||||: |||:
207 IAsnValIThrSerAsnLeuAsnGlnGlnIleGlyPheGlyAspA 224
|||||:|||||: |||:|||||: |||:|||||: |||:
743 AAAAAATTAACATAGCCCA.....TATGCT 768
|||||:|||||: |||:|||||: |||:|||||: |||:
224 snSerLysHisSerProGluLysLeuLysGluValLeuSerGlnAsn 240
769 TTTTACCAACAGAGGCTCATTTGGCGACAGTGGCTACCAATGTTTAT 818
|||||:|||||: |||:|||||: |||:|||||: |||:
241 AlaLeuThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAl 257
819 CTATGATGCCCAAAAGCAAAAGCGTTAATTAATGAGGCTATGCAACAG 868
|||||:|||||: |||:|||||: |||:|||||: |||:
257 aTyrAspLysGlnGlnLysArgTyrPalaPheLeuGlyAlaTyrAspTyr 274
|||||:|||||: |||:|||||: |||:|||||: |||:
869 GCAACCCCTATATAGGAAAGCAATGCTCCAGCTAGTGTGTAAGAT 918
|||||:|||||: |||:|||||: |||:|||||: |||:
274 rPalaGlyTyr.....GlnLysAsnSerTrpGln.....Glu 284
919 TGG...TTCTATGATGAAATCTTGTGCGAGATACCATTCAGTATCTA 965
|||||:|||||: |||:|||||: |||:|||||: |||:
285 TrpAsnIleTyrLysLysGlnPheAlaAspGluIleLysGln..... 298
|||||:|||||: |||:|||||: |||:|||||: |||:
966 CGAACCAATCAATGCAAAATGCAAAATCTTTTAAAGCAATTAATATGCG 1015
|||||:|||||: |||:|||||: |||:|||||: |||:
299 .....ArgAspAsn.....A 302
1016 CAGGAAATATGATGCCAAACATTAACACTATCTTACCTTATACATTA 1065
|||||:|||||: |||:|||||: |||:|||||: |||:
302 IArgIleThrIleLysGlyTyrGlyGlnHis.....Trp 313
|||||:|||||: |||:|||||: |||:|||||: |||:
1066 AAAACACGAAACCGTTCAATGTTTAAATGTTCTTATTCGAGACAGCANG 1115
|||||:|||||: |||:|||||: |||:|||||: |||:
314 LysThrThrGlyThr.....AsnSerHisIleGlySerThrAlaVa 327
|||||:|||||: |||:|||||: |||:|||||: |||:
1116 AGAACCTGTTTATCATGCTGCAGGTGGGTCAACAGTTATGACCCAGAC 1165
|||||:|||||: |||:|||||: |||:|||||: |||:
327 IArgLeuAlaGlyAsnGluArgGlyAla..... 336
1166 TGAATTAATGAGAAATATTTCTTATTTACAAAGAAAGTGAATG 1215
|||||:|||||: |||:|||||: |||:|||||: |||:
337 ..AsnAsnGlyGlnAsnValIThrPhe.....GluAsnAsnGlyThrLeu 350
|||||:|||||: |||:|||||: |||:|||||: |||:
1216 ATACTTACCAAGCAATCAACCAAGGCGGCGGTTGATTTTGAAGG 1265
|||||:|||||: |||:|||||: |||:|||||: |||:
351 ValLeuAspGlnAsnIleAsnGlnGlyAlaGlyLeuPhePheLysG 367
|||||:|||||: |||:|||||: |||:|||||: |||:
367 yAspTyrThrValLysGlyAlaAsnAsnGlyIleThrTrpLeuGlyAlaG 384
|||||:|||||: |||:|||||: |||:|||||: |||:
1313 GCGTTTCATATCAGTATGAGCAGATACCTTACTACTGGAAGTAAAGCGGTG 1362
|||||:|||||: |||:|||||: |||:|||||: |||:
384 IyIleAspValAlaAspGlyLysLysValValIleTrpGlnValLysAsnPro 400
|||||:|||||: |||:|||||: |||:|||||: |||:
1363 GCAACGACCGCTGTCCAAATGCGCAAGGCAAGCGACGCTGTGTTCAAGC 1412
|||||:|||||: |||:|||||: |||:|||||: |||:
401 AsnGlyAspArgLeuAlaLysIleGlyLysGlyThrLeuGluLysAsnG 417
|||||:|||||: |||:|||||: |||:|||||: |||:
1413 CAAAGGGAAACCAAGGCTCGGTACGCGTGGCGACGCTGAAGCTATCT 1462
|||||:|||||: |||:|||||: |||:|||||: |||:
417 yThrGlyValAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIle 434
|||||:|||||: |||:|||||: |||:|||||: |||:

```

1463 TAGATCAGCAGCGAGCATCAAGGCAAAAACAGCCTTTAGTGAATC 1512
|||||
434 euasnglInalaspAlaspIysValGlnalapheserGlnVal 450
1513 GCGTTGGCAGCGGAGCGAGCGTGCMAACTGATCCGATTAATCAGTT 1562
|||||
451 GilyleValSerGlyArgGlyThrLeuValLeuAsnSerProAspGlnI 467
1563 CAACCCGACAAACTCTATTTCGGCTTCGCGGCGAGCTTTGGATTGA 1612
|||||
467 eAsnProAsnAsnLeuThrPheGlyPheArgGlyGlyArgLeuAspAla 484
1613 ACGGGCATTCGCTTCCTCCACCGCATTAATAATACGATGAAGGGCG 1662
|||||
484 snGlyAsnAspLeuThrPheGlnHisIleArgAsnValAspGlnGlyAla 500
1663 ATGATTGTCAACCCACATCAAGCAAAAGATTCACCGCTACCATACAG 1712
|||||
501 ArgIleValAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrG 517
1713 CAATTAAGATTAATCTACA..... 1731
|||||
517 YlySerLeuIleThrAlaProGlnAsnLeuSerValTyrGlnIleArg 534
1731 1731
534 snAspTyrAspAspAspTyrTyrGlyTyrTyrSerTyrArgGlyPro 550
1731 1731
551 IleProGlnIlyLysAspLeuTyrTyrLysAsnTyrArgTyrAlaLe 567
1732ACCG 1735
567 ulysSerGlyLysValAsnAlaProMetProGluAsnGlyGlnThrG 584
1736 GCATTAAC..... 1743
|||||
584 IuAsnAsnAspTrpIleLeuMetGlySerThrGlnGluAlaLysLys 600
1744ACAACTTGGATTAGCAAAAAGAAATTCCTACAAAGCTTG 1784
|||||
601 AsnAlaMetAsnHisLysAsnAsnGlnArgIleSerGlyPheSerGly 617
1785 GTTGGCGGAAAGATGCAACCAAAAGCAAGCGCGCTCAATCTGAT 1834
|||||
617 ePheGlyGlnGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsn 634
1835 ACCAACCAGGAAGGAGGATCGCACTTCTGCTTCCGGCGAACAAT 1884
|||||
634 heAsnGlyLysSerAlaGlnAsnArgPheLeuLeuThrGlyThrAsn 650
1885 TTAACGGCAATATCAGCAAAACAAAGCAAACTTTTTCAGCGGGCG 1934
|||||
651 LeuAsnGlyLysIleSerValThrGlnGlyAsnValLeuLeuSerGly 667
1935 ACCGACACCGCAGCGCTACATCATTTAGGAAGCGGTGTGCAAAAATG 1984
|||||
667 gProThrProHisAlaArgAspPheValAsnLysSerSerAlaGlnLys 684
1985 AAGGT.....ATCCCAAGAGGAAGAAATCGTGGGACACAGATTCGATC 2028
|||||
684 sPAlaHisPheSerLysAsnAsnGluValValPheGlnAspAspTrpIle 700
2029 GACCGCACAATTAAGCGGAAACTTCATATT...CAGGGCGGCAAGC 2075
|||||
701 AsnArgThrPheLysAlaThrGluIleThrValAsnGlnSerAlaSer 717
2076 GGTGGTTCCGCAATGTGGCAAAAGTGAAGCGCATTCGCTTTAAACA 2125
|||||
717 eSerSerGlyArgAsnValSerAsnIleThrAlaAsnIleThrAlaThr 734

2126 ATCAGCCCCAAGCAGTTTCGGGTGTCACCGCATCAAGCCACACAAATC 2175
|||||
734 sPAsnAlaLysValAsnLeuGlyTyr.....LysAsnGlyAspGluVal 748
2176 TGTACAGCTTCGAGCTGAGCGGCTGACAGAGTTGATCAAGAAAACCAT 2225
|||||
749 CysValArgSerAspTyrThrGlyTyrValThrCysAsnThrGlyAsnLe 765
2226 TACCGCAGTAAGATGATTCCTCATTTGACGAAGCCGACATCAGAGCA 2275
|||||
765 user...AspLysAlaLeuAsnSerPheGlyAlaThrLysIleAsnGly 781
2276 ATGTCAACCTTCGCGATCAGCTCATTTTAATCTGACAGAGATTGGCCA 2325
|||||
781 snValAsnLeuAsnGlnAsnAlaIleValLeu...GlyLysAlaIle 796
2326 CTCACGCGCAATCTTAAGTCAGCGGAGACACGACATATACGTTACGCG 2375
|||||
797 LeuThrGlyGlnIleGlnGly..... 803
2376 CAACGCCACCAAAAGCGCAACCTCAGCCTCGGCAATGCGCCAGCAA 2425
|||||
804GlnGlyAsnSerArgVal 810
2426 CATTTAATCAGCCACATTAACGGCAACACATCGCTTCGACAAATGCT 2475
|||||
810 eLeuAsnGln..... 813
2476 TCATTTAATCTAAGCAACAGCCGCTACAAAACGCGAGTTCAGCCTTC 2525
|||||
813 813
2526 CGACAACGCTAAGGCAACGTAAGCATTCGCGACTCAAGGCAATGCT 2575
813 813
2576 CCTAGCCGATAGCAGATTCATTTGAAAACAGCGCTTACCGGA 2625
813 813
2626 AAATCAGCGGCGGCAAGATACGGCATTTACATTAAGACAGCGAATG 2675
|||||
814HisSerLysTr 817
2676 GACGTCGCGCGGCGGCAAGATTAAGCAATTTAACTTCAACAGCGCA 2725
|||||
817 PHisLeuThrGlyAspSerGlnValHisAsnLeuLeuAlaAspSer 834
2726 CCATTACATCAATTCGCTATGACAGCATGCGGAGCGCGCAAAAC 2775
|||||
834 IHisLeuLeuAsnAla.....SerAspAlaGlnSerAla..... 846
2776 GGCAGTGGCGAGATGCCCGCGCGCGCTTCGCGCGCTTCCTATTATC 2825
846 846
2826 CGTTAGCGCGGCAACTTCGCGAGAAATCCGTTTCAACAGCTGACGTA 2875
847AsnGlnTyrHisThrLeuLysIle 855
2876 ACGGCAATTAAGCGGTGAGGAAATTCGCTTATGTGCGAATCTTC 2925
|||||
855 sn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuAla 870
2926 GCGTAACGCGGCGCAATTAAGCTTCGCGGAAAGTTCCGAAGGCACTTA 2975
|||||
871 GluAsnLeuGlyAspLysValLeuValLysGlnSerIleSerGlyHisTyr 887
2976 CACCTTGCTGCTCAACAATACCGGCAAGCAACCGGTAATCTCGACAA 3025
|||||
887 rGlnLeuHisValGlnAspLysThrGlyLysPro...AsnGlnGlnGly 903
3026 TGACGGTAGTGGAAGAAAGACACACACCGGCTGTCCGAAATCTTAAT 3075


```
302 laGlyThrIleLysGlyTyrGlyGluHis.....Trp 313
1066 AAAACGACGAAACGGTCAATGTTTAAATGTTTATCCGACGACGAAG 1115
314 LysThrThrGlyThr.....AsnSerHisIleGlySerThrAlaVala 327
1116 AGAACCTGTTTATCATGCTGACGAGTGGGCTCAACAGTTATGACCCAGAC 1165
327 LArgLeuAlaGlyAsnGluArgGlyAla..... 336
1166 TCATTAATGAGAGAAAATATTTCTTATTGACAAAGAAAGTGAATG 1215
337 ..AsnAsnGlyGlnAsnValThrPhe.....GluAsnAsnGlyThrLeu 350
1216 ATACTTACGACACATACCAAGCGCGCGCGCTTGTATTTGAGAG 1265
351 ValLeuAspGlnAsnIleAsnGlnGlyAlaGlyLeuPhePheLysGly 367
1266 TAAATTTACGGTCTCGCTTAAACACAGAA...ACGTGCAAGCGCGG 1312
367 YAspThrThrValLysGlyAlaAsnAsnGlyLeThrThrLeuGlyAlaG 384
1313 GCGTTCATATCATGATGACGACGACCTTACTTGAAAGTAAACGCGTG 1362
384 LylLeAspValAlaAlaAspGlyLysValValThrPglValLysAsnPro 400
1363 GCAACGACCGCGCTGCCAATAATGGCAAGGACGCTGCTGTTCAAGC 1412
401 AsnGlnAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGly 417
1413 CAAAGGAGAAAAACAAGGCTCGGTGACGCTGCGACGCTAAAGTCATCT 1462
417 YThrGlyValAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleL 434
1463 TAGATCAGCAGCGCGACCATCAAGCAAAAAACAAGCTTATGTAATC 1512
434 euAsnGlnGlnAlaAlaAspLysLysValGlnAlaPheSerGlnVal 450
1513 GGCTGTGTCAGCGCAGCGGAGCGGTGACACTGAATGCCGATTAATCATTT 1562
451 GlyIleValSerGlyArgGlyThrLeuValLeuAsnSerProAspGlnI 467
1563 CAACCCGCAAAATCTATTTCCGCTTCCGCGGCGAGCTTTGATTTGA 1612
467 euAsnProAsnAsnLeuTyrPheGlyPheArgGlyGlyLysLeuAspAla 484
1613 ACGGCGATTTCCGTTCTGTTCCACCGCATTCAAAATACGATGAAGGCGG 1662
484 snGlyAsnAspLeuThrPheGlnHisIleArgAsnValAlaSpGlnGlyAla 500
1663 ATGATTGTCAACACACATCAAGCAAGAAAGATCCACGTTATCATTCAGG 1712
501 ArgIleValAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGly 517
1713 CAATTAAGATTTACTACA..... 1731
517 YLysSerLeuIleThrAlaProGlnAsnLeuSerValTyrGlnIleArg 534
1731 ..... 1731
534 snAspTyrAspAspAspAspTyrTyrGlyTyrTyrSerTyrArgLysPro 550
1731 ..... 1731
551 IleProGlnGlyLysAspLeuTyrTyrLysAsnTyrArgTyrTyrAlaLe 567
1732 .....ACCG 1735
567 YLysSerGlyLysSerValAsnAlaProMetProGlnAsnGlyGlnThrG 584
1736 GCAATTAAC..... 1743
584 LuAsnAsnAspTyrIleLeuMetGlySerThrGlnGluGluAlaLysLys 600
1744 .....AACACTGATAGCAAAAAAGAAATTCCTTACACGCTTG 1784
601 AsnAlaMetAsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPh 617
1785 GTTTGGCGAGAAAGATGCACCAACCAAAACGAGCGGCGCTCAATCTGATT 1834
617 ePheGlyGlnGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPro 634
1835 ACCAACCGGAGAAAGACGATGCGCACTTTACTGCTTTCCGCGGACAAAT 1884
634 heAsnGlyLysSerAlaGlnAsnArgPheLeuThrGlyGlyThrAsn 650
1885 TTAACCGCAATATCAACGCAAAACAGCGCAACTGTTTTCAGCGGACG 1934
651 LeuAsnGlyLysIleSerValThrGlnGlyAsnValLeuLeuSerGlyArg 667
1935 ACCGACACCGACGCGCTACATCATTTAGGACGCGGCTGCAAAATG 1984
667 gProThrProHisAlaArgAspPheValAsnLysSerSerAlaGlnLysA 684
1985 AAGGT.....ATCCCAAGAGAGAAATCGTGTGGCAACAGATGGATC 2028
684 spAlaHisPheSerLysAsnAsnGlnValValPheGluAspAspTyrPile 700
2029 GACCGCATTTTAAACGCGGAAACTTCCATATT...CAGGCGGACACAGC 2075
701 AsnArgThrPheLysAlaThrGlnIleThrValAlaSerIleSerAlaSerPh 717
2076 GGTGTTTCCCGCAATGTTGCCAAAGTGAAGGAGATGGCATTTAAGCA 2125
717 eSerSerGlyArgAsnValSerAsnIleThrAlaAsnIleThrAlaThrA 734
2126 ATCAAGCCCAAGCAGTTTTCGTGTCGACCGCATCAACGACACAAATC 2175
734 spAsnAlaLysValAsnLeuGlyTyr.....LysAsnGlyAspGluVal 748
2176 TGATCAAGCTTGGAGCTGACGGGTGACAAAGTGTACCAAAAAACCAAT 2225
749 CysValAlaPheSerAspTyrThrGlyTyrValThrCysAsnThrGlyAsnLe 765
2226 TACCGAGATTAAGTATGCTTCATTTAGACAGCAAGACGACATCAGAGCA 2275
765 user...AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyA 781
2276 ATGTCAAGCTTCCCGATCAGCTCATTTAAATCTCACAGAGACTTGGCACA 2325
781 snValAsnLeuAsnGlnAsnAlaAlaLeuValLeu...GlyLysAlaAla 796
2326 CTCACAGCGCAATCTTAGTGACGCGGAGACAGCACATTAACGTTACGGG 2375
797 LeuThrPglGlnIleGlnGly..... 803
2376 CAACGCCACCCAAACGCGCAACCTCAGCCTGTGGCAATGCCCAAGCA 2425
804 .....GlnGlyAsnSerArgValS 810
2426 CATTTAATCAAGCACATTAAACGCAACACATCGGCTTCGCAATGCT 2475
810 erLeuAsnGln..... 813
2476 TCATTTAATCTAAGCAACACGCGGTACAAACGCGCAGTGTACGCTTTC 2525
813 ..... 813
2526 CGAACACGCTTAAGCAACGTAAGCATTCGCACTCAACGCGCAATGCT 2575
813 ..... 813
2576 CCTAGCGGATTAAGCAGTATTCATTTTGAACACAGCGGCTTTACCGGA 2625
813 ..... 813
```

2626 AAAAAACAGCGGCGGAGATACGCACTTAAAGACAGCAATG 2675
 814HisSerLysTr 817
 2676 GAGCGTCGCGGCGGAGATAGCAATTTAACTTACACAGCGCA 2725
 817 PHisLeuThrGlyAspSerGlnValHisAsnLeuSerIleAlaSpSerH 834
 2726 CCATTACACATCAATCCGCTATGACACAGATGGCGGCGGCAACC 2775
 834 IHisLeuHisAsnAla.....SerAspAlaGlnSerAla..... 846
 2776 GGCAGTCGGCGGAGATGCCCGCGCGCTGCGCGCTCCATTATC 2825
 846 846
 2826 CGTTACGCGCGCACTTCGGAGATCCGTTTACACAGCGTAA 2875
 847AsnGlnTyrHisThrLeuLysIleA 855
 2876 ACGGCAATTGACGTCAGGAGACATCCGCTTATGTCGCACTTTC 2925
 855 sp...HisLeuSerGlyAsnGlnHisPheHisTyrLeuThrHisLeuAla 870
 2926 GCGTACCGCAGCGCAATTTGACGTCGGCGGAAAGTCCGAGGCACTTA 2975
 871 GlnAsnLeuGlyAspLysValLeuValLysGlnSerAlaSerGlnHisTyr 887
 2976 CACCTTGCTGTACACATPACCGGAGAACCGCTAGTCTCGAGCAAT 3025
 887 GlnLeuHisValGlnAspLysThrGlyLysPro...AsnGlnGlyLys 903
 3026 TGACGCTAGTCGAGCAAAACACACACACCGCTCCGCAAAATCTTAAT 3075
 903 LeuAsnLeuPheAspAlaSerValAlaGlyAspArgSer...HisLeuSer 918
 3076 TTCACCTTCGCAAAACGACGTCGATCCGCGCGCATGGCTTATCAGCT 3125
 919 ValSerLeuAlaAsnHisHisValAspLeuGlyAlaLeuArgTyrThr 935
 3126 TATCGCGCAAGCGCGGAGTCCGCTGCTATTAATCCGCTCAAGACAG 3175
 935 eLysThrGlnAsnGlyIleThrArgLeuTyrAsnProTyrAlaGlnAla 952
 3176 AGCTTCGCAAACTCGGCAAGCGGAGAAACAGAGCGGCTTGAAG 3225
 952 Arg.....ArgArgValLysProAlaProSerProAlaThrAsnThr 965
 3226 GCAAAACAGGCACTTCGCGCGCAACACAGCGGGAAGAACAC 3273
 966 AlaSerGlnAlaGlnThrAspSerAlaGlnIleAlaLysProGlnAsn 981
 seq_name: sp_bacteria:088093
 seq_documentation_block:
 ID 088093 PRELIMINARY; PRT; 1377 AA.
 AC 088093;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HAEMOGLOBIN PROTEASE PRECURSOR.
 GN HBP.
 OS Escherichia coli.
 OG Plasmid pColy-K30.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98416173; PubMed=9743528;
 RA Otto B.R., van Dooren S.J.M., Nuijens J.H., Luijck J., Oudega B.,
 RT "Characterization of a hemoglobin protease secreted by the pathogenic
 Escherichia coli strain Ebl.*";

RL J. Exp. Med. 188:1091-1103(1998).
 DR EMBL; AJ223631; CAA11507.1; ..
 DR MEROPS; S06.003; ..
 DR InterPro; IPR00710; IGA_S6.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF02395; IGAL; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 KW Hydrolase; Plasmid; Protease; Serine protease; Signal.
 FT SIGNAL 1 52 POTENTIAL.
 SQ SEQUENCE 1377 AA; 148256 MW; BB16D898EDAC0416 CRC64;
 alignment_scores:
 Quality: 1126.50 Length: 1530
 Ratio: 1.375 Gaps: 64
 Percent Similarity: 53.529 Percent Identity: 26.144
 alignment_block:
 US-09-303-518D-653 x 088093 ..
 Align seg 1/1 to: 088093 from: 1 to: 1377
 82 TTAGCCATATGCGCTGTCGCTTTCGCGATTTGCCCCAGCGCGGAGACA 131
 42 IleProValLeuPheSerAlaGlySerLeu.....AlaGlyThr 54
 132 CACTTATTTTGGCATCACTACCAATACATATGCGGACTTTGCCGAAATA 181
 54 ValAsnAsnGlnLeuGlyTyrGlnLeuPheArgAspPheAlaGlnAla 71
 182 AAGCGCAATTTGCGATGCGGCGGCAAAAGATATGAGTTTACCAACAAAA 231
 71 LysGlyMetPheArgProGlyAlaThrAsnIleAlaIleLysAsnLysGln 87
 232 GGGGAGTGTGTGCGCAATATGATGACGAAACCGCGCATTTGATTTTTC 281
 88 GlnLeuPheValGly...ThrLeuAspLysAlaAlaMetProAspPhe 103
 103 ValValAlaSerGlnIleGlyValAlaThrLeuIleAsnProGlnTyr 120
 282 TGTGTA...TCGCTAACCGCGTGGCGGCGCATTTGGCGGAGATATA 328
 329 TTGTGACGTCGACATACGCGGCTATACATGTTGATTTTGGTGG 378
 120 LeuAspSerValLysHisAsnGlyGlyTyrThrAsnValSerPheGly 135
 379 GAGGGAAGCAATCCGATCGACACCGCTTCTTACCAATTTGAAAG 428
 136 AspGlyGlnAsn.....ArgTyrAsnIleValAlaAspArg 146
 429 AATTAATTATTAAGCAGGAGCTAACGCGCATCTTATGGCGGATATTC 478
 146 GAsnAsnAlaProSer.....LeuAspPheH 155
 479 ATATGCGCGCTTTCACAAATTTGTACAGATGCACAGCTGTGGAGATG 528
 155 IAlaProArgLeuAspLysLeuValThrGlnValAlaProThrAlaVal 171
 529 ACCAGTAT...ATGATGGGTGGAATACCGCATTAATTAATTAATACC 575
 172 ThrAlaGlnGlyAlaValAlaGlyAlaTyrLeuAspLysGlnTyrTyr 188
 576 TGATCGTGTGCAATGCGAGCAGACAGCAATTTGCGGCTGTGATGAG 625
 188 ovalPheTyrArgLeuGlySerGlyThrGlnTyrIleLysAspSerAsnG 205
 626 ACGAACCAATTAACCGGGAAGTTATATCATATTCAGAACGCAATTTCT 675
 205 LysLeuThrLys.....MetGlyGlyAlaTyrSer 215
 676 TGGTGTGTCGTCGCAATACCTTTGCA.....CAAATGCA... 711
 216 ThrLeuThrGlyGlyThrValGlySerLeuSerTyrTyrIleAsnGly 232

```
712 .....TCAGTGTGGCAGACATCACTTAGTAGCGAAAAATTA 751
      |||
232 uMetIleSerThrSerGlyLeuVal.....PheA 243
752 AACATAGCCCATGTGTTTATACCAACGAGCGCTCATTTGGCAGACT 801
      |||
243 sPyrTrpIleuAsnGlyAlaMetProIleTyrGlyLunIaGlyAspSer 259
802 GGCTCCACCAATGTTTATCATATGATGCCCAAAAGCAAAAGTGTATTAA 851
      |||
260 GlySerProLeuPheAlaPheAspThrValGlnAsnLysTyrValLeuVal 276
852 TGGGATTATGCAACAGCAACCCCTATATAGAAAAAACAATGGCTTCC 901
      |||
276 IGIValLeuThrAlaGlyAsnGlyAlaGlyLysIleValArgLysAsnTyrP 293
902 AGCTACTGTGTAAGATGGTCTCATGATGAATCTTGGTCGAGATACC 951
      |||
293 IAlValIleProLeuAspPheIleGlyGlnLysPheAsnGlnAspAsn 309
952 CATTCAGTATCTACGACACCATCAAAATGG.....AAATACTTTT 995
      |||
310 AlaProValThrPheArgThrSerGlyGlyAlaLeuGluTyrSerPh 326
996 TAACGACAATTAATATGGCGCAGAAAAATCGATGCCAAACATAAACACT 1045
      |||
326 eAsn...SerSerThrGlyAlaGlyAlaLeuThrGlnGlyThrThrThr 342
1046 ATTCTCTACCTTATAGATTAAAAACAGCAACCGCTTCAATGTTTATGTT 1095
      |||
342 YrAlaMet..... 344
1096 TCTTATCCGAGACAGCAAGAACCTGTTTATCATGCTGAGTGGGT 1145
      |||
345 .....HisGlyGlnGlyAs 350
1146 CAACAGTATGACCCAGACTGATTAATGAGAAAAATATTCCTTATATG 1195
      |||
350 nAsp.....LeuAsnAlaGlyLysAsnLeuIlePheGlnG 362
1196 ACAAAAGAAAGGTGAATGATCTTACCAGCAACATCAACCAAGCGCG 1245
      |||
362 IyGln...AsnGlyGlnIleAsnLeuLysAspSerValSerGlnIyAla 377
1246 GCGGTTGTATTTAGGGTAATTTTACGCTGCTGCTTAAACAAACGA 1295
      |||
378 GlySerLeuThrPheArgAspAsnTyrThrValThrThrSerAsnGlySe 394
1296 AACGTGGCAAGCGCGCGGCTTCATATACATGATGACAGTACCGTTACTT 1345
      |||
394 rThrTrpThrGlyAlaGlyIleValValAspAsnGlyValSerValAsnT 411
1346 GGAAGATTAACGGCGTGGCAACGCGCTGTCCAAATTCGGCAAAAGC 1395
      |||
411 rGlnIleValAsnGlyValLysGlyAspAsnLeuHisLysIleGlyGlnGly 427
1396 ACGTGTGTGTTCAAGCCAAAGGGAACCAAGGCTGCTGACGCTGG 1445
      |||
428 ThrLeuThrValGlnGlyThrGlyLysAsnGlnGlyGlyLeuLysValG 444
1446 CGACGGTAAAGTCATATGATCAGCAGCGCGGAGCATGAAGCAAAAAAC 1495
      |||
444 yAspGlyLysValValLeuAsnGlnGlnAlaAspAsnLysGlyGlnValG 461
1496 AAGCTTTAGTAATCGGCTGTGTCAGCGGCGGAGCGGAGCGTGAACATG 1545
      |||
461 IAlaIlePheSerSerValAsnIleAlaSerGlyTyrProThrValValLeu 477
1546 AATGCCGATTAATCACTCAACCCGACAAACTATTTGGCTTTCCGGG 1595
      |||
478 ThrAspGluArgGlnValAsnProAspThrValSerTyrGlyTyrArgG 494

1596 CGACGTTTGATTTGAACGGGCGATTCGCTTGGTCCACCGCATTCAA 1645
      |||
494 YGIYThrIleuAspValAsnGlyAsnSerLeuThrPheHisGlnIleuLysA 511
1646 ATACCGATGAAGGGCGATGATGATTCACCAACATCAACAGCAAAAGATCC 1695
      |||
511 IAlaIleAspTyrGlyAlaValLeuAlaAsn...AsnValAspLysArgAla 526
1696 ACCGTTACCATTTACA..... 1710
      |||
527 ThrIleThrLeuAspTyrAlaLeuArgAlaAspLysValAlaLeuAsnG 543
1711 ....GGCAATTAAGATATTACTACACCGGCAAT.....AACA 1744
543 YTrpSerIleuSerIleuLysGlyThrAlaGlyAsnLeuTyrLysTyrAsn 560
1745 ACAACTTGGATAGCAAA.....AAGAAATTCCTTAC 1776
560 snProTyrThrAsnThrThrAspTyrPheIleuLysGlnSerThrTyr 576
1777 AACGTTGGTT..... 1788
      |||
577 ...GlyTyrPheProThrAspGlnSerSerAsnAlaThrTyrGluPheVal 592
1789 .....GGCGAAGAGATGCAACCAA..... 1809
592 IGIHisSerGlnGlyAspAlaGlnLysLeuValAlaAspArgPheAsnT 609
1810 .....ACGACGGGCGGCTCAATCTG... 1830
      |||
609 hrAlaGlyTyrLeuPheHisGlyGlnLeuLysGlyAsnLeuAsnValAsp 625
1831 AATTACCAACCGGAGAAAGCGGATCGCATTTTCTGCTCCGCGGAC 1880
      |||
626 AsnArgLeuProGlnGlyValThrGlyAlaLeuAlaMetAspGlyAlaAl 642
1881 AAATTTAACGCGCAATATCACGCAACAAACGCGCAACTGTTTTCAGCG 1930
      |||
642 aAspIleSerGlyThrPheThrGlnGlnAsnGlyArgLeuThrLeuGlnG 659
1931 GCACAGCGACCGCGCATCATCAT..... 1959
      |||
659 IynIleProValIleHisAlaTyrAsnThrGlnSerValAlaAspLysLeu 675
1960 TTAGACGCGGCTGTCAAAATGAGATATCCACAGAGCAAGAAATCGT 2009
      |||
676 AlaIleSerGlyAspHisSerVal.....LeuThrGlnProThrSerPh 690
2010 GTGGACAGCATGTGATCGACCGCATTTTAAAGCGAAACTTCATTA 2059
      |||
690 eSerGlnGlnAspTyrPrlGlnAsnArgSerPheThrPheAspArgLeuSer 707
2060 TTCAGGGCGCAACAGCGGTGTTCCCGCAATGTT.....GCCAA 2100
      |||
707 eulysAsnThrAspPheGlyLeuGlyArgAsnAlaThrLeuAsnThrThr 723
2101 GTGACAGCGCATTTGGCATTTAAGCAATCAGCGCCCAACGATTTGCGTGT 2150
      |||
724 IleGlnAlaAsp.....AsnSerSerValThrLeuGly... 734
2151 CGCACGCGCATCAAGCCACACATCTGTATACAGCTTGGAGTGGACGGGTC 2200
      |||
735 .....AspSerArgValPheIleAspLysAsnAspGlyGlnGly... 747
2201 TGACAATGTGTACCGAAAAACCATTAACGACATTAAGATGTTGCTTCA 2250
      |||
748 .....ThrAlaPheThrLeuGlnGlnGlyThrSerValAlaThr 760
2251 .....TTGACGAAGACCGCATCAGAGCAATGTCAGCTTCCGATCA 2294
      |||
761 LysAspAlaAspLysSerValPheAsnGlyThrValAsnLeuAspAsnG 777
2295 CGCTCATTTAATTCACAGGACTTGCACACATCAACGCGCAATCTTAGTG 2344
```

```

777 nserValIleuasnIleasnAspIle.....PheasnIlyGlyIleGlnA 792
2345 CAGCGGAGACACGCACTATACGTTACCGCAACGCCCAACCAAC... 2391
792 la.....AsnAsnSerThrValAsnIle 799
2392 ...GGCAACCTGACGCTGGGCAATGGCCAGACACATTTAATTCAGC 2438
800 SerSerAspSerAlaValIleuIysnSerThrLeuThrSerThrAlaLe 816
2439 CACATTTAAACGGCAACACATCGGCTTCGAGAAATGCTTCATTTAATCTAA 2488
816 uasnIleuasnIlysglyAlaAsnAlaLeuAlaSerIleSerPhe...ValS 832
2489 GCAACACAGCCCGCTACAAAACGGCACTGACGCTTCC..... 2526
832 erAspGlyProValAsnIleSerAspAlaThrLeuSerLeuAsnSerArg 848
2526 ..... 2526
849 ProAspGluValSerHisThrLeuProValIlyAspTyrAlaGlySe 865
2527 .....GACACGCTAAGCAACGTAAGC...CATTCGG 2557
865 rTrpAsnLeuIysGlyAspAspAlaArgLeuAsnValGlyProTyrSerM 882
2558 CACTACAGCGCAATGCTCCCTACCGCGATAG..... 2589
882 etLeuSerIlyAsnIleAsnValGlnAspIlyThrValThrLeuGly 898
2589 ..... 2589
899 GlyGluGlyLeuLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTy 915
2590 .GCAGTATTCATTTTGAACACGCGCTTACCGGAAATACAGCGCGC 2638
915 rSerLeuPheAsnGlyTyrArgAsnIleTrpSerIlySerLeuAsn...A 931
2639 GCAAGGATTCGGCATTTACACTAAAGACAGGATGAGCGTCCGCGC 2688
931 laProAspAlaThrValSerMetThrAspThrGlnTrpSerMetAsnIly 947
2689 GGCACGGAATTAGCAATTTAAACCTGACAGCGCACCATTTACACTCA 2738
948 AsnSerThrAlaGlyAsnMetIlyLeuAsnArgThrIleValGlyPheAs 964
2739 TTCGCCCTATCGACACGATGCGGACGCGCAACCGGACGTGCGCAG 2788
964 n..... 964
2789 ATGCGCGCGCGCGCGCTTCGCGCGCTCCATATATCCGTTACGCGGCA 2838
964 ..... 964
2839 ACTTCGGCAGAAATCCGTTTCAACACGCTGACGTAACGGCAATTTGA 2888
965 ...GlyGlyThrSerProPheThrThrLeuThrThrAspAsnLeuAspAl 980
2889 CGGTACGGAGACATTCGCTTATGTGCGAACTTCGCGTACCGCAGC 2938
980 aValGlnSerAlaPheValMetArgThrAspLeu.....AsnIlyAlaA 995
2939 GCAATTTAGAGCTGGCGAAGATCCGGAAGC...ACTTACACCTTGCT 2985
995 sPllySerLeuValIleAsnIlySerAlaThrGlyHisAspAsnSerIleTrp 1011
2986 GTCAC.....AATACCGGCAACGAAACCC... 3009
1012 ValAsnPheLeuIlySerProSerAsnIlyAspThrLeuAspIleProIle 1028
3010 .GTAGTCTCGAGCAATTGAC..... 3030

```

```

1028 uValSerAlaProGluAlaThrAlaAspAsnLeuPheArgAlaSerThrA 1045
3031 .GTACGTGAGAGAAAAGAACACACCGCTGTCGCAAAATCTTAATTTC 3078
1045 rValValGlyPheSerAspValThrProIleLeuSerValArgIlyGlu 1061
3079 ACCCTGCAAAACGACACACGTCGATGCGCGCATGCGCTTATCAGCTTAT 3128
1062 AspGlyIlyIysGluTrpValLeuAspGly.....TyrGlnValAl 1075
3129 CCGCAAGACGCGGAG.....TTCGCGC 3151
1075 aArgAsnAspGlyGlnGlyAlaAlaThrPheMetHisIleSerT 1092
3152 TGCAATATCGGTCAAAGAA.....CAAGAG 3177
1092 TyrAsnAsnPheIleThrGluValAlaAsnLeuAsnIlySarMetGlyAsp 1108
3178 CTTTCGACAACTCGGCAAGCGGAGAAACAGAGCGCGCTTGACGCGC 3227
1109 LeuArgAspIleAsnGlyAlaIleGlyThrTrpValArgLeuAsnGln 1125
3228 AAACACAGCA.....CAAC 3241
1125 ySerGlySerAlaAspGlyIlyPheThrAspHisTyrThrLeuLeuGlnM 1142
3242 TTGCGCGCGCAACAAACAGCGGAAAGACACCGCGCAAGCTTGACGCG 3291
1142 etGlyAlaAspArgIlyIysHisGlyLeuGlySerMetAspLeuPheThrGly 1158
3292 CTGATTCGCGCGCGCGCAATGCCACCGCAAAAGCAGAAAGTGTGCCGA 3341
1159 ValMetAlaThrTyrThrAspThrAspAlaSerAlaAspLeuTyrSerG 1175
3342 ACCGCGCGCGCAGCAGCGCGGAAATGCCGCGCATTTATGACGCGGAG 3391
1175 IlyThrIlySerThrIlyGly.....GlyPheTyrAlaSerGlyL 1189
3392 AAGCAAAACAGGTCGACGCGGATTAAGACACCGCTTGCGGAAACAG 3441
1189 eupPheArgSerGlyAlaTyrPheAsp.....ValIleAlaIlyTyr 1202
3442 CGCGAAGCGGAACCGCGCGCTACACCGCTTCCCGCGCGCGCGC 3491
1203 IleHisAsnGluAsnIly..... 1208
3492 CGCGCGCGGGAATTCGCGCAACCGGACGCCCAACCGCAACCGCAG 3541
1208 ..... 1208
3542 AGCGGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAATTTTCC 3591
1209 ..TyrAspLeu.....AsnPheAlaGlyAlaGlyIlyGlnAsnPheArg 1222
3592 GCCACGCTCAACAGCGTTTCCCGCTACAGAGAGAA..... 3627
1223 Ser.....HisSerLeuTyrAlaGlyAlaGlyValGlyTyrArgTyrH 1237
3628 ...TTGACGCGCGCTTTCGCAAGACGCGCAACGCGCTTTCGACAA 3673
1237 sLeuThrAspThrThrPheValGluProGlnAlaGluLeuValTrp... 1252
3674 GCGGATTCGCGGACACCAACTACCGTTGCAAGATTTCCGCGCTTAC 3723
1253 .....GlyArgLeuGlnIlyGlnThrPheAsnTrpAsn 1263
3724 GCGCAACAAACGACCTGCGCAATGCGTAATGCAAGAAACCTCGGCGAG 3773
1264 AspSerGlyMetAsp.....ValSerMetArgArgAsnSerValAs 1277
3774 C.....GGGCGCTGCGCATCTGTTTGCACAAACGCGACGCGAA 3814
1277 nProLeuValGlyIlyArgThrGlyValVal.....SerGlyL 1289

```


833 spgLyProValAsnIleSerAspAlaIleuSerLeuAsnSerArgPro 849
2506AACGGACGCT 2516
850 AspGluValSerHisThrLeuLeuProValTyrAspTyrAlaGlySerTr 866
2517 GACGGCT...TCCGACAAAGCTTAAGCAACGTAAG...CATTCGGCAG 2560
866 pAsnLeuysGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetL 883
2561 TCACGGCAATGTCCTCCAGCCGATAG..... 2589
883 euserGlyAsnIleAsnValGlnAspLysGlyThrValThrLeuGlyGly 899
2590GC 2591
900 GluGlyGluLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrSe 916
2592 AGTATTCATTTTGAAACACGCGCTTACCGGAAATACAGCGCGCA 2641
916 rLeuPheAsnGlyTyrArgAsnIleTrpSerGlySerLeuAsn...Alap 932
2642 AGGATACGCAATTACATTAAAGACAGCAATGAGCGCTCCGCGGCG 2691
932 rAspAlaThrValSerMetThrAspThrGlnTrpSerMetAsnGlyAsn 948
2692 ACGGAATTAGCAATTAACCTTGACAAACGCCACCATACACCAATTC 2741
949 SerThrAlaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsn... 964
2742 CGCTATGACACAGATGCGGCGGCAACCGCGCAGTCGCGAGATG 2791
964 964
2792 CGCGCGCGCGCTTCGCGCGCTTCCTATTATCCGTCAGCGCCAACT 2841
964 964
2842 TCGGAGATCCCGTTTCACACGCTGACGGGAACGGCAATGGAACG 2891
965 GlyGlyThrSerProPheThrThrLeuThrThrAspAsnLeuAspAlaVa 981
2892 TCAGGACATTCGCTTTATGTCCGAATCTTCGCTACGCGGCA 2941
981 GlnSerAlaPheValMetArgThrAspLeu....AsnLysAlaAspL 996
2942 AATTGAGCTGGCGAAGTCCGAAGC...ACTTACACCTGGCTGTC 2988
996 yLeuValIleAsnLysSerAlaThrGlyHisAspAsnSerIleTrpVal 1012
2989 AAC.....AATACGGCAACGAACCC...GT 3011
1013 AsnPheLeuLysLysProSerAsnLysAspThrLeuAspIleProLeuVa 1029
3012 AAGTCTCGACCAATTGACG.....G 3031
1029 lSerAlaProGluAlaThrAlaAspAsnLeuPheArgAlaSerThrArgV 1046
3032 TGTGTGAGGAAAGACACACACGCGCTGTCGAAATCTTAATTCAC 3081
1046 AlValGlyPheSerAspValThrProIleLeuSerValArgLysGluAsp 1062
3082 CTGCAAAACGACACGTCGATGCGCGCATGCGCTTACCTATGCTCG 3131
1063 GlyLysLysGlyTrpValLeuAspGly.....TyrGlnValAlaVal 1076
3132 CAAGAAGCGCGAG.....TTCGCGCTGC 3154
1076 gAsnAspGlyGlnGlyLysAlaAlaAlaThrPheMetHisIleSerTyrA 1093
3155 ATAAATCGGTCAAAGA.....CAAGAGCT 3180
1093 snAsnPheIleThrGluValAsnAsnLeuAsnLysArgMetGlyAspLeu 1109

3181 TCCGACAAATCTCGCAAGCGGAGAGAAAGACAGCCGCTTGACGGCA 3230
1110 ArgAspIleAsnGlyGluAlaGlyThrTrpValArgLeuLeuAsnLysSe 1126
3231 ACAGCA.....CAACTTG 3244
1126 rGlySerAlaAspGlyGlyPheThrAspHisThrThrLeuLeuGlnMetG 1143
3245 CCGCAACACACAGCGGGAAGAAAGACAGCGCAAGCCGTGACGGCTG 3294
1143 LysAlaAspArgLysHisGlnLeuGlySerMetAspLeuPheThrGlyVal 1159
3295 ATTGCGCGCGCGGCAATGCCACCGAAGGCAAGAGTGTGCGCAAC 3344
1160 MetAlaThrTyrThrAspThrAspAlaSerAlaAspLeuTyrSerGlyL 1176
3345 GCGCGCGGACAGCGCGGGAATAAGCCGATTAAGCAGCGGAGAGAG 3394
1176 sThrLysSerTrpGlyGly.....GlyPheTyrAlaSerGlyLeuP 1190
3395 AGAAAAACGGGTGCAGCGGATTAAGACACCGCCTTGCGGAACAGCG 3444
1190 heArgSerGlyAlaTyrPheAsp.....ValIleAlaLysTyrIle 1203
3445 GAAAGCAACCGCGCGGCTACACCGCCTTCGCCGCGCGCGCGCG 3494
1204 HisAsnGluAsnLys..... 1208
3495 CCGCGGAGATTTCGCGCAACCGACCGCCCAACCGCAACCGCAAC 3544
1209T 1209
3545 GCGACCTGATCAGCGCTTATCCCAATAGCGTTGATGAATTTCCGC 3594
1209 yAspLeu....AsnPheAlaGlyAlaGlyLysGlnAsnPheArgSer 1223
3595 ACGCTCAACAGCGTTTCGCGCTACAGGACGA..... 3627
1224HisSerLeuTyrAlaGlyAlaGlyValGlyTyrArgTyrHisLe 1238
3628 .TTGACCGCGGTGTTTCCGAAGACGCGCGCAACGCCGTTTGACAAAG 3676
1238 uThrAspThrThrPheValGluProGlnAlaGlyLeuValTrp..... 1252
3677 GCATCCGGGACACCAACTACCGTTCGCAAGATTCGCGGCTACCGC 3726
1253GlyArgLeuGlnGlyIleThrPheAsnTrpAsnAsp 1264
3727 CAACAAACGACGCTGCGCAATCGGTATGCAAGAAACCTCGGACG... 3774
1265 SerGlyMetAsp.....ValSerMetArgArgAsnSerValAsnPr 1278
3775GGCGCGCTCGGATCCTGTTTGCACACCGGACCGGAACA 3817
1278 OlLeuValGlyArgThrGlyValAla.....SerGlyLysT 1290
3818 CCTTC.....GACGACGGCATCGGCAACTCGGACGCGCTGGCCAC 3858
1290 hrPheSerGlyLysAspTrpSerLeuThrAlaArgAlaGlyLeuHisTyr 1306
3859 GGTGCGCTTTTCGGCAATACGGCATCGGAGGTTGACATCGGACATCAG 3908
1307GluPheAspLeuThrAspSe 1313
3909 CGCGGCGCGGCTTTAGTACGCGACCTTTCAGACGCGATCAGAGGCA 3958
1313 rAlaAspValHisLeuLysAspAlaAlaGlyGlnHisGlnIleAsnLysA 1330
3959 AATTCGCGCGCGCGCTGCTGCAT...TACGCAATTACGCAAGATAC... 4002
1330 rGlyAspSerArgMetLeuTyrGlyValGlyLeuAsnAlaThrPheGly 1346


```

1153 TATCGACACGACGTGAATATGAGAAAAATATTCCTTATGACAAAG 1202
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 .....LeuAsnAlaGlyGlnSerLeuValPheSerGlyGln.. 361
1203 AAAAGGTGAATGATTAATACGAGACATCAACAGCGCGCGGCTT 1252
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 .AsnGlyAlaIleIleLeuLysSerValThrGlnGlyAlaGlyTyrL 378
1253 TGTATTTTGAAGGTATTTTACGCTCTGCGCCCTAAAAACAGCAACGTCG 1302
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
378 euGlnPheLysSerTyrProValSerAlaGlnSerGlyLysThrTrp 394
1303 CAAGCGCGCGGCTTCATCATGATGACGATGACGCTTACTTGGAAAGT 1352
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
395 ThrGlyAlaGlyIleIleThrAspLysGlyThrAsnValThrTrpLysVa 411
1353 AAACGGCGTGGCAAAAGCAGCCGCTGCCAAATCGGCAAAAGCAGCGTCG 1402
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
411 .AsnGlyValAlaGlyAspAsnLeuHisLysLeuGlyGlnGlyThrLeu 428
1403 TGGTTCAACCAAGGGGAAACCAAGGCTCGTCAGCGCTGGCGACGCT 1452
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
428 hrIleAsnGlyThrGlyValAsnProGlyGlyLeuLysThrGlyAspGly 444
1453 AAAGTCATCTTGAATCAGCAGCGGACGATCAAGGCAAAACAAAGCCTT 1502
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 .IleValValLeuAsnGlnGlnAlaAspThrAlaGlyAsnIleGlnAlaAs 461
1503 TACTGAATCGCGCTTGGTCAGCGGCGAGGGGAGCGTGCATGTAATGCGG 1552
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
461 eSerSerValAsnLeuAlaSerGlyArgProThrValAlaLeuGlyAspA 478
1553 ATATAGAGTTCAACCCGCAAACTCTATTTGGCTTTCGCGGCGAGCT 1602
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
478 .IaArgGlnValAsnProAspAsnIleSerTrpGlyTyrArgGlyLys 494
1603 TTGGATTTGAACGGCGATTCGCTTTCCTCCACCGCATTCAAATATGCGCA 1652
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 .LeuAspLeuAsnGlyAsnAlaValThrPheThrArgLeuGlnAlaAlaAs 511
1653 TGAAGGGCGGATGATTCACACCAATCAAGCAAAAGAA..... 1692
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 pTyrGlyAlaValIleThrAsnAsnAlaGlnGlnLysSerGlnLeuLeuL 528
1693 .....TCCACGCTTACCATTAACGAGC 1713
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
528 euAspLeuLysAlaGlnAspThrAsnValSerGlnProThrIleGlyAsn 544
1714 AATAAGATATTACTACAAACCGCAATTAACACACTTGGATAGCAAA.. 1761
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 .IleSerProPheGlyThrGlyThrProGlyAsnLeuLysSerMetIle 561
1762 .....AAGAATATTCCTTACACAGC 1780
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 eLeuAsnSerGlnThrArgPheTyrIleLeuLysSerAlaSerTyrGlyA 578
1781 GT.....TGG..... 1785
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
578 snThrLeuTrpGlyAsnSerLeuAsnAspProAlaGlnTrpGlnPheVal 594
1786 ....TTTGGCGAGAAAGATGCACAAACAGC..... 1812
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
595 GlyMetAsnLysAsnLysAlaValGlnThrValLysAspArgIleLeuAl 611
1813 .....ACGGCGCGC 1822
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
611 agIlyArgAlaLysGlnProValIlePheHisGlyGlnLeuThrGlyAsn 628
1823 TCATCTGAATTAACAAACGAGAAAGCGATGCACTTACTGCTTTC 1872
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
628 eLysPheAlaIleProGlnValProGlyGlyArgLysValIlePheAsp 644
1873 GCGGCAACAAATTTA...AAGCGATATATCAGCGCAACAAAGCGCAACT 1919
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
645 GlySerValAsnLeuProGlnGlyThrLeuSerGlnAspSerGlyThrIle 661
1920 GTTTTTCAGCGCGGACGACCGGACCGGCTACATCATTTAGCAAGCG 1969
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 uIlePheGlnGlyHisProValIleHisAla...SerIleSerGlySerA 677
1970 GGGTGTCAAAATGAGAGTATCCCAAGAGAGAAATCTGTGGGACAAAC 2019
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
677 .IaProValSerLeuAsn.....GlnLys 684
2020 GATTGATTCGACCGCACATTTTAAAGCGAAACTTCATATTCAGGCGG 2069
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
685 AspTrpGlnAsnArgGlnPheThrMetLysThrLeuSerLeuLys.... 699
2070 ACAAGCGGTGTTCCCGCAATGTTGCCAAAGTGAAGCGGATTTGGCAT 2119
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
700 .....AspAlaAspPheHisL 705
2120 TTAGCAATCAGCCCAAGCAGTTCGCTGTCGACCGCATCAAGCCAC 2169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
705 euSerArgAsnAlaSerLeuAsnSerIleLysSerIaAspAsnSerHis 721
2170 .....ACAATCTGTACACGTTTCGACGTCGACGCGG 2198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
722 .IleThrLeuGlySerAspArgAlaPheValAspLysAsnAspGlyThrG 738
2199 TCTGCAAGTGTGTACCGAAAA.....ACCATTCGCG 2230
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
738 .YAsnTyrValIleProGlnGlnGlyThrSerValProAspThrValAsnA 755
2231 AGCATTAAGTATTCCTTCAATGACAGCAACCGCATCAGAGCGCAATGTC 2280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
755 sp.....ArgSerGlnTyrGlnGlyAsnIle 763
2281 AGCCTTGGCGATCAGCGCTATTAATCTC.....ACAGCA.. 2316
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
764 ThrLeuAsnHisAsnSerAlaLeuAspIleGlySerArgPheThrGlyG 780
2316 ..... 2316
780 yIleAspAlaTyrAspSerAlaValSerIleThrSerProAspValLeuL 797
2316 ..... 2316
797 euThrAlaProGlyAlaPheAlaGlySerSerLeuThrValHisAspGly 813
2317 .....CTTGGCACCTCAGCGCAATCTTACTGACGCGGACAGCAGCA 2360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 GlyHisLeuThrAlaLeuAsnGlyLeuPheSer.....AspGlyHis 827
2361 CTATACGGTTACGGCGCACCGCACCGCAACCGCACTCAGCGCGG 2410
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
827 stIle.....GlnAlaGlyLysAsnGlyLysIleThrLeuSerG 840
2411 GCAT...GCCCAAGCAACATTTATCAAGCCACA..... 2442
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
840 .LYThrProValLysAspThrAlaAsnGlnTyrAlaProAlaValTyrLeu 856
2443 TTAACGCGCACACATCGGCTTCGACAAATGCTTCAATTAATCAAGCA 2492
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
857 ThrAspGlyTyrAspLeuThrGlyAspAsnAlaAlaLeuGlnIleThrAr 873
2493 CAAGCGC.....GTAC 2503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
873 gGlyAlaHisAlaSerGlyAspIleHisAlaSerAlaAlaSerThrValT 890
2504 AAACGCGACGTCGCTTCGCAACAGCTTAAAGCGAAACGTAAGCGAT 2553
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
890 hrIleGlySerAspThrProAlaGlnLeuAlaSerAlaGlnThrAlaAla 906
2554 TCCGCACTCAGCGCAATGTCCTCCAGCGCATTAAGCGATATTCATTT 2603

```



```

seq_name: sp_bacteria:Q9AL58
seq_documentation_block:
ID Q9AL58 PRELIMINARY; PRT: 1373 AA.
AC Q9AL58;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PIC.
GN Shigella flexneri 2a.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=42897;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YSH6000T;
RX MEDLINE-20231776; PubMed-10768931;
RA Al-Hasani K., Henderson I.R., Sakellaris H., Rajakumar K., Grant T.,
  Natario J.P., Robins-Browne R., Adler B.;
RT "The sigma gene which is borne on the she pathogenicity island of
  Shigella flexneri 2a encodes an exported cytoplasmic protease involved
  in intestinal fluid accumulation.";
RL Infect. Immun. 68:2457-2463(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-YSH6000T;
RX MEDLINE-21105956; PubMed-11162180;
RA Al-Hasani K., Rajakumar K., Bulach D., Robins-Browne R., Adler B.,
  Sakellaris H.;
RT "Genetic organization of the she pathogenicity island in Shigella
  flexneri 2a.";
RL Microb. Pathog. 30:1-8(2001).
DR EMBL; AF200692; AAK00464.1;
DR InterPro; IPR00710; IGA_56.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
SQ SEQUENCE 1373 AA; 146548 MW; 45C016CDAA577763 CRC64;

alignment_scores:
  Quality: 1056.00 Length: 1601
  Ratio: 1.344 Gaps: 72
Percent Similarity: 49.094 Percent Identity: 26.109

alignment_block:
US-09-303-518D-653 x Q9AL58 ..
Align seg 1/1 to: Q9AL58 from: 1 to: 1373

60 CATCCGCTTCTGCGCCGCTTACTTACCATATGCGCTTGGCGCATC 109
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
37 Hs1leuenua1a1y1le..Pro1a1le1cys1eu1cys1tyr1ser1 53
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
110 TGCCCAAGCCGCGGCGGACACACTTATTCGCGATCACTACCAATAC 159
   : ||| ||||| ||||| ||||| ||||| ||||| |||||
53 lesergin.....Alag1yle1val1ar1ser1as1ple1al1ar1 67
   ||| ||||| ||||| ||||| ||||| ||||| |||||
160 TATCGACTTTCGCGAAATTAAGCAAGTTTGCAGTGGGGGCAAGA 209
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 Tyr1Ar1sp1he1a1gl1u1as1n1y1s1gl1eu1p1he1val1p1ro1 84
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 TATGAGGTTTACACAAAAGAGGAGTGTGCGCAATGATGATACGA 259
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 p1le1pro1val1tyr1asp1y1s1p1e1ly1s1eu1val1gl1y1ar1 100
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 AACCCCGATGATTGTTTCTGTGTATCGCGTAACGCGGTGGCGGCA 309
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
100 ysa1la1p1ro1me1ta1a1sp1he1ser1Ser1Val1ser1as1n1y1a1a1thr 116
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
310 TTGGCGGCGCATATATATGTGACGCGGCGCATACGCGGCTATTA 359
   ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

117 Leu1val1ser1pro1gl1u1tyr1le1val1ser1val1y1s1a1s1n1gl1y1tyr1 133
360 CAATGTTGATTTCGTCGCGGAGGAGCAATCCCGATACACACCGCTTT 409
   : ||||| ||||| ||||| ||||| ||||| ||||| |||||
133 gSer1val1Ser1he1Gly.....Asn1gl1y1s1a1s1n.....T 143
410 CTTACCAAAATTCGAAAGAAATATATTAAGCAGGACTAACGGCCAT 459
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 htyr1ser1leu1val1asp1arg1asn1s.....His 152
460 CCTTATGCGCGGCGATTATCATATGCCCGCTTGCACAAATTTGCACA 509
   ||| : ||||| ||||| ||||| ||||| ||||| |||||
153 Pro...Ser1le1as1n1gl1y1a1a1p1ro1gl1eu1as1n1y1s1eu1val1thr1 168
510 TGCAGAACCTGTGATGATGACCACTTATATGATGGGTGAATATACGCT 559
   : ||||| ||||| ||||| ||||| ||||| |||||
168 uval1le1pro1ser1a1val1thr1Ser.....Glu1gl1y1Thr1ys...Ala 182
560 ATTAAATTAATATACCTGATCGT.....GTTGCAATCGGAGCA 597
   : ||||| ||||| ||||| ||||| ||||| |||||
182 sna1a1tyr1lys1tyr1Thr1Glu1arg1tyr1Thr1a1phe1tyr1Arg1Val1Gly1Ser 198
598 GCGACACATATTCGCGGCTGATGAGAGCAACCCATAACCGCGAAG 647
   ||| ||||| ||||| ||||| ||||| ||||| |||||
199 Gly1Thr1Gln1tyr1Thr1ys...Asp1Lys1Asp.....Gly1as 209
648 TTCATATCATATATGCAAGCGCATATCTTGGCGTGCAGCAATACCT 697
   : ||||| ||||| ||||| ||||| ||||| |||||
209 n1eu1val1y1s1Val1a1a1gl1y1tyr1a1a1phe1y1Thr1Gly1tyr1Thr. 225
698 TTGCACAAAATGATCAGGTGGTGGCACAGTCACTTAGTAGCGAATA 747
   226 .....Gly1Val1Pro1leu1le1Ser1Asp1a1Thr 234
748 ATTAACATATACCCA.....TATGTTTTTATAC 776
   ||| : ||||| ||||| ||||| ||||| |||||
235 lle1val1Ser1as1n1pro1gl1u1tyr1Thr1tyr1as1n1Pro1val1as1n1gl1y1Pro1leu1 251
777 AACAGGAGCTCATTTGGCGAGCTGCGTCCACCAATGTTATCATATG 826
   | ||||| ||||| ||||| ||||| ||||| |||||
251 oAsp1tyr1Gly1a1a1Pro1gl1y1Asp1Ser1Gly1Ser1Pro1leu1p1he1a1tyr1asp 268
827 CCAAAAGCAAAAGTGTTAATTAATGGGTATTGCAAAACAGCAACCC 876
   ||||| ||||| ||||| ||||| ||||| |||||
268 yS1n1Gln1Lys1tyr1Val1lle1Val1a1a1Val1leu1Arg1a1tyr1a1Gly 284
877 TATATAGGAAAAGCAATGCGCTCCAGCTAGTGTAAAGATGGTCTTA 926
   ||||| ||||| ||||| ||||| ||||| |||||
285 lle1as1n1gl1y1a1a1Thr1as1n1TP1r1p1asn1val1le1Pro1thr1asp...Tyr1le 300
927 TGATGAATCTTTGCTGAGATATCCATTCAGTATTCACGACACCATC 976
   : ||||| ||||| ||||| ||||| ||||| |||||
300 uas1n1Gln1Val1met1Gln1asp1asp.....Phe1asp1a1a1Pro1Val 313
977 AA.....AATGGAAATCTTTTAAACGAC 1002
   : ||||| ||||| ||||| ||||| ||||| |||||
313 sp1he1Val1ser1Gly1Leu1n1gl1y1Pro1leu1as1n1tyr1Thr1.....Asp 326
1003 AATATATATGCGCGCAAAAATATGATGCCAAACATTAACATATTCCT 1052
   : ||||| ||||| ||||| ||||| ||||| |||||
327 Lys1Thr1Ser1Gly1Thr1Gly1Thr1Leu1ser1Gln1y1Ser1Lys1as1n1TP1r1p1he 343
1053 ACCTATATGATTAATAAACACGACCGTTCATTTGTTAATGTTCTTAT 1102
   : ||||| ||||| ||||| ||||| ||||| |||||
343 L..... 343
1103 CCGAGACAGCAAGAGAACCTGTTATTCATGCTGCAGGTGGGCTCAACACT 1152
   ||||| ||||| ||||| ||||| ||||| |||||
344 .....His1Gly1Gln1y1s1Asp1asn1asp... 350
1153 TATGACCCAGACATGATATATGAGAAAATATTCCTTATATGACAAAG 1202
   ||||| ||||| ||||| ||||| ||||| |||||
351 .....Leu1as1n1a1a1Gly1Gln1s1n1Leu1val1Phe1ser1Gly1Gln... 362

```

```

1203 AAAAGTGAATTGATCTTACACGAAACATCAACGAGCGCGCGGT 1252
1204 ::::: ::::: ::::: ::::: ::::: :::::
1205 .AsnglyAlaIleIleuLysAspSerValThrGlnGlyAlaGlyTyrL 379
1206 ::::: ::::: ::::: ::::: ::::: :::::
1207 TGTATTTGAGGGTAAATTTTACGGTCTCGCTAAACAAACGAAACGTG 1302
1208 ::::: ::::: ::::: ::::: ::::: :::::
1209 euGlubheYAspSerTyrProValSerAlaGluSerGlyLysThrTrp 395
1210 ::::: ::::: ::::: ::::: ::::: :::::
1211 CAAGGGCGCGCGCTTCATATCAGTGTGAGTACCGTTCATCTTGAAGT 1352
1212 ::::: ::::: ::::: ::::: ::::: :::::
1213 ThrGlyAlaGlyIleIleThrAspLysGlyThrAsnValThrTrpLysVa 412
1214 ::::: ::::: ::::: ::::: ::::: :::::
1215 AAAGCGCTGGCGAAACGACCGCTGCAAAATCGGCAAGGACGAGCTGC 1402
1216 ::::: ::::: ::::: ::::: ::::: :::::
1217 LAsnGlyAlaGlyAspAsnLeuHisLysLeuGlyGlnGlyThrLeuT 429
1218 ::::: ::::: ::::: ::::: ::::: :::::
1219 TGGTTCAGCCAAAGGGGAAACAGGCTGCGTCAAGCTGGGCGACGCT 1452
1220 ::::: ::::: ::::: ::::: ::::: :::::
1221 hLLeAsnGlyThrGlyValAsnProGlyGlyLeuLysThrGlyAspGly 445
1222 ::::: ::::: ::::: ::::: ::::: :::::
1223 AAAGTATCTGTGATCAGCAGCGCGGATCAAGGCAAAACAAACGCTT 1502
1224 ::::: ::::: ::::: ::::: ::::: :::::
1225 LLeValValLeuAsnGlnGlnAlaAspThrAlaGlyAsnLLeGlnAlaPh 462
1226 ::::: ::::: ::::: ::::: ::::: :::::
1227 TAGTGAATCGGCTTGGTCAGCGGAGGAGGAGGAGTCAACTGAATGCGG 1552
1228 ::::: ::::: ::::: ::::: ::::: :::::
1229 eSerSerValAsnLeuAlaSerGlyAArgProThrValValLeuGlyAspA 479
1230 ::::: ::::: ::::: ::::: ::::: :::::
1231 ATATATCAGTTCACCCGCAACACTCTATTTGCGCTTTCGCGCGGAGCT 1602
1232 ::::: ::::: ::::: ::::: ::::: :::::
1233 LArgGlnValAsnProAspAsnLLeSerTrpGlyTyrArgGlyLys 495
1234 ::::: ::::: ::::: ::::: ::::: :::::
1235 TTGATTTACGCGGATTCGCTTCGTTCCGCGCATTCATAAATACCGA 1652
1236 ::::: ::::: ::::: ::::: ::::: :::::
1237 LeuAspLeuAsnGlyAsnAlaValThrPheThrArgLeuGlnAlaAlaAs 512
1238 ::::: ::::: ::::: ::::: ::::: :::::
1239 TGAAGGGCGGATGATTGTCAACCAACATCAACAAAGAA 1692
1240 ::::: ::::: ::::: ::::: ::::: :::::
1241 pTyrGlyAlaValIleThrAsnAsnAlaGlnGlnLysSerGlnLeuL 529
1242 ::::: ::::: ::::: ::::: ::::: :::::
1243 ::::: ::::: ::::: ::::: ::::: :::::
1244 TCCACCGTTACATTCACAGGC 1713
1245 ::::: ::::: ::::: ::::: ::::: :::::
1246 euAspLeuLysAlaGlnAspThrAsnValSerGlnProThrLLeGlysn 545
1247 ::::: ::::: ::::: ::::: ::::: :::::
1248 AATAAGATATTACTACACCGGCAATTAACAACTTGATAGCAAA 1761
1249 ::::: ::::: ::::: ::::: ::::: :::::
1250 ILeSerProPheGlyGlyThrGlyThrProGlyAsnLeuTyrSerMetI 562
1251 ::::: ::::: ::::: ::::: ::::: :::::
1252 ::::: ::::: ::::: ::::: ::::: :::::
1253 AAAGAAATTCCTACACG 1780
1254 ::::: ::::: ::::: ::::: ::::: :::::
1255 eLeuAsnSerGlnThrArgPheTyrIleLeuLysSerAlaSerTyrGlyA 579
1256 ::::: ::::: ::::: ::::: ::::: :::::
1257 GTr 1785
1258 ::::: ::::: ::::: ::::: ::::: :::::
1259 suThrLeuTrpLysAsnSerLeuAsnAspProAlaGlnTrpGluPheVal 595
1260 ::::: ::::: ::::: ::::: ::::: :::::
1261 TTTGGCGAGAAAGATGCAACCAAG 1812
1262 ::::: ::::: ::::: ::::: ::::: :::::
1263 GlyMetAsnLysAsnLysAlaValGlnThrValLysAspArgIleLeuAl 612
1264 ::::: ::::: ::::: ::::: ::::: :::::
1265 ::::: ::::: ::::: ::::: ::::: :::::
1266 AACGGCGCGC 1822
1267 ::::: ::::: ::::: ::::: ::::: :::::
1268 agLysrGlnLysGlnProValIlePheHisGlyGlnLeuThrGlyAsnM 629
1269 ::::: ::::: ::::: ::::: ::::: :::::
1270 TCAATCTCAATTTACCAACCGGAGAGGATCGCACTTACTGCTTTC 1872
1271 ::::: ::::: ::::: ::::: ::::: :::::
1272 eAspValAlaIleProGlnValProGlyGlyArgLysValIlePheAsp 645
1273 ::::: ::::: ::::: ::::: ::::: :::::
1274 GGGCAACAAATTTA 1919
1275 ::::: ::::: ::::: ::::: ::::: :::::
1276 GlySerValAsnLeuProGlnGlyThrLeuSerGlnAspSerGlyThrLe 662
1920 GTTTTTCAGCGGACAGCCGACACCGGACCGCTACATCATTTAGAACG 1969
1921 ::::: ::::: ::::: ::::: ::::: :::::
1922 uLLePheGlnGlyHisProValIleHisAla 1969
1923 ::::: ::::: ::::: ::::: ::::: :::::
1924 SerLeuSerGlySerA 678
1925 ::::: ::::: ::::: ::::: ::::: :::::
1926 GGTGTCAAAATGGAAGGTATCCACAGAGAGAAATCTGTGGGACAC 2019
1927 ::::: ::::: ::::: ::::: ::::: :::::
1928 LArgProValSerLeuAsn 685
1929 ::::: ::::: ::::: ::::: ::::: :::::
1930 GATTGGATTCAGCGCACATTTAAAGCGGAAACTTCATATTCAGGCGG 2069
1931 ::::: ::::: ::::: ::::: ::::: :::::
1932 AspTrpGlnAsnArgGlnPheThrMetLysThrLeuSerLeuLys 700
1933 ::::: ::::: ::::: ::::: ::::: :::::
1934 ACAAGCGGTGTTCGCCCAATGTGCCAAAGTGAAGCGGATTCGATT 2119
1935 ::::: ::::: ::::: ::::: ::::: :::::
1936 AspAlaAspPheHisL 706
1937 ::::: ::::: ::::: ::::: ::::: :::::
1938 TAAACCAATCAGCCCAAGAGATTTCGGTTCGACCGCATCAAGGAC 2169
1939 ::::: ::::: ::::: ::::: ::::: :::::
1940 euSerArgAsnAlaSerLeuAsnSerAspLysSerAspAsnSerHis 722
1941 ::::: ::::: ::::: ::::: ::::: :::::
1942 ACATCTGTACACGTTGCGACTGGACGG 2198
1943 ::::: ::::: ::::: ::::: ::::: :::::
1944 LLeThrLeuGlySerAspArgAlaPheValAspLysAsnAspGlyThrG 739
1945 ::::: ::::: ::::: ::::: ::::: :::::
1946 TCTGACAAAGTTGTACGAGAAA 2230
1947 ::::: ::::: ::::: ::::: ::::: :::::
1948 YAsnTyrValIleProGlnGlyThrSerValProAspThrValAsnA 756
1949 ::::: ::::: ::::: ::::: ::::: :::::
1950 AGCATAAAGTATGCTTCATTGAGCAAGACCGCATCAGAGGCAATGTC 2280
1951 ::::: ::::: ::::: ::::: ::::: :::::
1952 ArgSerGlnGlyGlyAsnLLe 764
1953 ::::: ::::: ::::: ::::: ::::: :::::
1954 AGCTTCGCCATCAGCTCATTTAATCTC 2316
1955 ::::: ::::: ::::: ::::: ::::: :::::
1956 ThrLeuAsnHisAsnSerAlaLeuAspLLeGlySerArgPheThrGly 781
1957 ::::: ::::: ::::: ::::: ::::: :::::
1958 YLLeAspAlaTyrAspSerAlaValSerIleThrSerProAspValLeu 798
1959 ::::: ::::: ::::: ::::: ::::: :::::
1960 euThrAlaProGlyAlaPheAlaGlySerSerLeuThrValHisAspGly 814
1961 ::::: ::::: ::::: ::::: ::::: :::::
1962 CTGCCACACTCAGCGCAATCTTAAGTCAGCGGAGACACGCA 2360
1963 ::::: ::::: ::::: ::::: ::::: :::::
1964 GlnHisLeuThrAlaLeuAsnGlyLeuPheSer 828
1965 ::::: ::::: ::::: ::::: ::::: :::::
1966 AspGlyHis 828
1967 ::::: ::::: ::::: ::::: ::::: :::::
1968 CTATACGGTTACCGGACGACCGCAACGAGGCAACCTCAGCTCGTGG 2410
1969 ::::: ::::: ::::: ::::: ::::: :::::
1970 GlnAlaGlyLysAsnGlyLysIleThrLeuSerG 841
1971 ::::: ::::: ::::: ::::: ::::: :::::
1972 GCCCAAGCACTTAATCAAGCACA 2442
1973 ::::: ::::: ::::: ::::: ::::: :::::
1974 LysThrProValLysAspThrAlaAsnGlnTyrAlaProAlaValTyrLeu 857
1975 ::::: ::::: ::::: ::::: ::::: :::::
1976 TTAAGCGGCAACACATCGGCTTCGACAAATGCTTCATTAATCAAGCA 2492
1977 ::::: ::::: ::::: ::::: ::::: :::::
1978 ThrAspGlyTyrAspLeuThrGlyAspAsnAlaAlaLeuGlnLLeThrAr 874
1979 ::::: ::::: ::::: ::::: ::::: :::::
1980 CAAAGCC 2503
1981 ::::: ::::: ::::: ::::: ::::: :::::
1982 gGlyAlaHisAlaSerGlyAspLLeHisAlaSerAlaAlaSerThrValT 891
1983 ::::: ::::: ::::: ::::: ::::: :::::
1984 AAAAGCGCACTGACGCTTTCGACACAGCTAAGGCAACGTAAGCCAT 2553
1985 ::::: ::::: ::::: ::::: ::::: :::::
1986 hLLeGlySerAspThrProAlaGlnLeuAlaSerAlaGlnThrAlaAla 907
1987 ::::: ::::: ::::: ::::: ::::: :::::
1988 TCCGCACTCAGCGCAATGCTCCCTAGCGGATAGGCAAGTATTCATTT 2603
1989 ::::: ::::: ::::: ::::: ::::: :::::
1990 SerAlaPheAlaGly 917
1991 ::::: ::::: ::::: ::::: ::::: :::::
1992 TGAACAGCGCGCTTACCGGAAATATCAGCGGCGGCAAG 2650

```


ID O9LA54 PRELIMINARY; PRT: 1335 AA.
AC O9LA54
DT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE EAC.
GN EAC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOR-9;
RX MEDLINE=20187497; PubMed=10722621;
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
RT activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214(2000).
DR EMBL; AF151674; AAF63038.1;
DR MEROPS; S06.0PW;-;
DR InterPro; IPR000710; IPA_S6.
DR Pfam; PF02395; IGAL; 1.
DR PRINTS; PRO0921; IGASERPTASE.
SQ SEQUENCE 1335 AA; 141672 MW; DC3BB4E6A4C969 CRC64;

alignment_scores:

Quality: 1027.50 Length: 1522
Ratio: 1.297 Gaps: 65
Percent Similarity: 52.037 Percent Identity: 26.478

alignment_block:

US-09-303-518D-653 x O9LA54

Align seg 1/1 to: O9LA54 from: 1 to: 1335

```

70 TGGCCCGCTTACTTAGCCATATGCTGCTGGCATTTGCCCAAGC 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44 SerGlyAlaValAlaGlyAlaLeuPheSerYrProSerLeuAlaSerVa 60
   : : : : : : : : : : : : : : : : : : : : :
120 CCGGGGGGACACACTTATTTGGCATCACTACCAATACATATGCGAGT 169
   : : : : : : : : : : : : : : : : : : : : :
60 L...ValGlyAsnThr.....LeuProTrpGlnThrYrAspGsp 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 TTGCGCAAAATGAAGCAAGTTTGCAGTCGGCGGCAAAATATGAGGTT 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
73 heAlaGlyAsnGlySglYrAlaPheHisAlaGlyAlaThrAsnIleProleu 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 TACACAAAAGGAGGAGTGTGGTGGCAATCGATGACGAAAGCCCGCAT 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 TyrGlyAlaGlyAsnGly...AlaValGlyGlyAlaGlyLeuAspLysAlaProMe 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 GATTGATTTTCTGTGTATCGCT...AACGGCGTGGCGCATTTGGCGG 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 tMetAspPheSerValValAspGlnIleLeuGlyAlaIleThrLeuIleS 122
   : : : : : : : : : : : : : : : : : : : : :
317 GCGATCATATATTTGTAGCGCTGGCAGCAATACGGCGCTATACAAAGTT 366
   : : : : : : : : : : : : : : : : : : : : :
122 erProGlnTyrValAlaGlyAlaValLysHisAsnGlySerYrAsnThrVal 138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 GATTGTGGTGGGAGGAGCAATCCGATGACGACCGCTTTCTTACCA 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 ArgPheGlyTyrAlaAspAspThr.....ThrTyrAr 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 AATTGTGAAAGAAATATATTAAGCAGGAGCTAACGGCATCTCTATG 466
   : : : : : : : : : : : : : : : : : : : : :
149 GlyLeuValAspArgAsnGlu.....HisT 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
467 GCGGCGATTAATGATGCGCGGTTTGCAACAATTTGTACAGATGAGAA 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 rPArgAspPheHisThrProArgLeuAsnGlyLeuValThrGlnValAla 173

```

```

517 CCGTTGAGATGACCAAGTATATGATGG...TGAAATATCGCTGATT 563
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174 ProValSerValThrAspAlaGlyThrGlyLysGlyValYrGlnAsnAr 190
   : : : : : : : : : : : : : : : : : : : : :
564 AAATTAATATCCGTGATCGTGGTGAATCGAGCAGCAGCAATATATGGC 613
   : : : : : : : : : : : : : : : : : : : : :
190 gSerArgTyrProValPheYrArgMetGlySerGlyThrGlnTyrThrG 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
614 GGTGATGATGAGAGCAAGCAACCCATACCGGAAAGTTATATCATATGCA 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
207 LyAlaAlaSerGly.....AlaLeuThrArgIleAla 217
   : : : : : : : : : : : : : : : : : : : : :
664 AGCCGATATTCCTGCTGCTGCGGCAATACCTTTGCAACAAATGATC 713
   : : : : : : : : : : : : : : : : : : : : :
218 GlyAlaTyrAlaTrpLysThrGlyGlyThrValGlySerProleuIleS 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
714 AGGTGGTGGCAGCTACCTAGCTAGCGGAAATTAATACATAGCCCAT 763
   : : : : : : : : : : : : : : : : : : : : :
234 rAspTrpSerLeuValSer...AsnProGlyTyrLeuTyrGlnSerValA 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
764 ATGTTTATTTACCAAGAGAGCTATTTGGCAGTGGCTGACCAATG 813
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 snGlyProleuAlaSerTyrGlyThrProLysPserGlySerProleu 266
   : : : : : : : : : : : : : : : : : : : : :
814 TTATCTATGATGCCCAAAAGCAAAAGTGTATTAATGAGGATATGCA 863
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 PheAlaTrpAspAlaValLysLysGlnTrpValLeuValAlaValleu 282
   : : : : : : : : : : : : : : : : : : : : :
864 AACAGCAACCCCTATATAGGAAAGCAATGGCTTCAGCTAGTTCGTA 913
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 .....AsnGlyTyrAlaGlyGlyLysGly.....LysT 292
   : : : : : : : : : : : : : : : : : : : : :
914 AAGATGTTCTATGATGAATATCTTCTGGAGATACCATCATGATATC 963
   : : : : : : : : : : : : : : : : : : : : :
292 hrasTrpPhe...ThrValIleProAlaGlyLysPValAsnAsnThrIle 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
964 TACGACACCATCA.....AATGGAAATA 989
   : : : : : : : : : : : : : : : : : : : : :
308 LysGlnAspSerSerGlyThrValAlaProAlaValAlaGlyLysPrl 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
990 CTTTTTAAC...GACAAATTAATGGCGGCAAGAAATCGATGCCAATC 1036
   : : : : : : : : : : : : : : : : : : : : :
324 eValTrpAsnTyrSerLysGlySerGlyGlyGlyThrLeuSerGlnAspG 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1037 ATAACACTATTCCTCTATGATATAGATTAATAACAGAACCTGTCAATG 1086
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
341 LysValThrLysMet..... 346
   : : : : : : : : : : : : : : : : : : : : :
1087 TTTAATGTTCTTATTCGAGACAGACAGAACTGTATTCATGCTGC 1136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
346 ..... 346
   : : : : : : : : : : : : : : : : : : : : :
1137 AGTGGGTCACAGATTATGCA...CCGAGACTGATTAATGGAGAAATA 1183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 .....AsnGlyPheArgGlyLysLeuAsnAspGlyLysAspI 360
   : : : : : : : : : : : : : : : : : : : : :
1184 TTTCCTTATTTGACAAAGAAAGTGAATGTACTTACCAACAGCAATC 1233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 leThrPhe.....GlyGlyLysGlyThrValValLeuLysAspVal 374
   : : : : : : : : : : : : : : : : : : : : :
1234 AACCAAGCGCGGCGGTTGTATTTTGAAGGTAATTTACGCTCGCC 1283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375 ValGlnGlyAlaGlySerLeuThrPheAsnGlyAspTyrThrValArgPr 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1284 TAAACAAACAGAAAGTGGCAAGCGCGGCGTCTCATATCAGTATGCA 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 OGluGlyAsnGlnThrTrpValGlyGlyIleIleValAsnAspGlyH 408
   : : : : : : : : : : : : : : : : : : : : :
1334 GTACCGTTACTTGGAAAGTAAAGCGGCTGGCAACAGACCGCTGTCCAA 1383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 isArgValAspTrpMetValAsnGlyLeuAlaGlyAspAlaLeuHisLys 424
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1384 ATCGGCAAGGACCGCTGCTGTTCAAGCCAAAGGGGAAACCAAGGCTC 1433

```



```

3144 GTTCGGCTGCATATCCGGTCAAGAACAGAGCTTTCGGACAACCTCG 3193
      ||| :||| ||| :||| :||| :||| :||| :||| :||| :|||
933 yPheSerHisIleThrProValValArgIaIaGluGlnGly..... 946
3194 GCAGGCGGGAGAAACAGAGCGCGCTTACAGCGCAAAACAGACGACTT 3243
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
947 ....ThrIlyIlyThrIaIaTrpGlnLeuAsnValValGlnProGluThr 961
3244 GCCGCCAACAACAGCGGAAAAAGACAGCGCAAGCCCTTGACGGCT 3293
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
962 ValAlaGlnProValValAspGluValSerArgProSerLeuProVal 978
3294 GATT.....GCGCGCGGGCGCAATGCCAAGCAAGCAAGCAAGTG 3334
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
978 lMetArgIaIaAspAlaIaIyThrProAsnProValGlnAlaValSerPro 995
3335 TTGCGCAACGGCGCGCGCGAGCGAGC.....GCGGAA 3366
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
995 aIProSerProAlaProAlaIaIaSerValThrAspValProThrIlyGlu 1011
3367 AATGCCGCAATTATGACAGCGGAGAGAAAGAAAAAGGGTGCAGCGGA 3416
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1012 AsnThrGly..... 1014
3417 TAAAGACACCGCTTGGCGAAACAGCGCAAGCGGAGCGCGCTA 3466
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1015 .....GluSerIlyGluTrpArgAspGluT 1023
3467 CGACCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCG 3516
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1023 hr..... 1023
3517 CAGCCCAACCGCAACCCCAACCGAGCGCGCATGATCAGCGCTTATGC 3566
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1024 .....ArgTrpLeuLeuThrGlyTyr.. 1030
3567 CAATAGCGGTGAGTGAAATTTCCGCCAAGCGTCAACAGCGTTTTCGCCG 3616
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1030 ..... 1030
3617 TACAGAGCAATTTGACCGCGGTGTTGCCGAAGACCGCGCAACCGCTT 3666
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1031 .....ArgSerThrVal 1034
3667 TGGACAAGCGGCATCCGGAGACCAACACTACCTGCGCAAGATTTCG 3716
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1035 AsnValSerAlaValArgAspAlaIyMetLeuMetSerMetGlyHisArg 1051
3717 CGCCTTAC.....CGCCACAACCGCACTGC 3742
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1051 gAsnPheIleAsnGluValAsnAsnLeuAsnLysArgMetGlyAspLeuA 1068
3743 GCCAATGGGTATGCAAAAAACCTC.....GGCAGC 3774
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1068 rGAspIleAsnLysGluIaIaGlyAlaTrpAlaArgIleMetSerGlyThr 1084
3775 GGGCGCGTGGCATCTGTTTCGACAAACGAGACCGCAACACCTTGA 3824
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1085 GlySerAlaIyGlyGlyPheSerAspAsnTrpThrHisValGlnValGln 1101
3825 C.....GACGCGATCGGCAACTCGGACGCGCTTG 3853
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1101 yAlaAspLysLysHisGluLeuAspGlyLeuAspLeuPheThrGlyValT 1118
3854 CCCAGCGCGCGCTTTCGGCAATACGCGATCGCGAGCTTGAC..... 3897
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1118 hrMetThrTyrThrAspSerAsnAlaGlySerAspTrpPheSerGlyLys 1134
3898 .....ATCGCATCAGCGCGCGCGGTGTTTAGTAGGG 3932
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1135 ThrLysSerValGlyAlaGlyLeuTrpAlaSerAlaLeuPheAspSerGln 1151

```

```

3933 CAGCCTTTCAAGCGGCATCAGAGCAAAATCCGCCGCCCGCTGCATT 3982
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1151 yAlaTyrIleAspLeuIle...GlyLys.....TyrValHisH 1163
3983 ACGCATTCAGGCAAGATATCCGCGAGCTTTCGGGATTCGGCATCGAA 4032
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1163 is.....AspAsnGluTyrThrAlaThrPheAlaGlyLeuGlyThrLys 1177
4033 CCGCAATCGGCGCAACGGCTATTGCTCCAAAACGGATTTCCATA 4082
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1178 .....AspTyr..... 1179
4083 CGAAACGTCATATTCGCCACCCCGCGCTTCATTCACCGGACCGCG 4132
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1180 .....GlySerHisSerTrpTyrA 1186
4133 CGGCATTTAGGCAGATTTATTCATTCACACCGCGCAACATTTCCATC 4182
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1186 lAcIyAlaGlnValGlyTyrArgTrpHisValThrGlnAspAlaTrpIle 1202
4183 ACGCCTTATTGAGCTGTCTCTATACGATGCGCGCTTCGGCAAGTCCG 4232
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1203 GluProGlnAlaGlnLeuValTyr..GlyAlaValSerGlyLysGlnPh 1218
4233 AACG.....CGGTCATATACCGCGCATTTGGCGCAGATTTTCG 4270
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1218 eSerTrpLysAspGlnGlyMetAsnLeuThrMetLysAspLysAspPheA 1235
4271 GCNAACCGCGAGTGGGAAATGGGCGTAAACCGCAATCAAGGTTTC 4320
      :||| :||| :||| :||| :||| :||| :||| :||| :|||
1235 snProLeuIleGly.ArgThrGlyValAspValGlyLysSerPhe..... 1249
4321 ACGGTGTCCTCCACGCTGCGCGCGCAAGGGCGCGCAATTTGAAGCGCA 4370
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1250 .....SerGlyLysAspTrpLys...V 1256
4371 GCACAGCGCGGCGCA 4384
      ||| :||| :||| :||| :||| :||| :||| :||| :|||
1256 alThrAlaArgAla 1260

seq_name: sp_bacteria:Q9LA58
seq_documentation_block:
ID Q9LA58 PRELIMINARY; PRT; 1335 AA.
AC Q9LA58;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE EAAA.
DE EAAA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOR-9;
RC MEDLINE=20187497; PubMed=10722621;
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
RT activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214(2000).
DR EMBL: AF151091; AAF63237.1;
DR MEROPS: S06.0PW; -.
DR InterPro: IPR000710; IqA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR SEQUENCE 1335 AA; 141677 MW; B7876C3536F7411 CRC64;

```

```

alignment_scores:
Quality: 1020.50      Length: 1506
Ratio: 1.293          Gaps: 67

```


Percent Similarity: 52.390 Percent Identity: 26.892

alignment_block:

US-09-303-518d-653 x Q9LA58 ..

Align seg 1/1 to: Q9LA58 from: 1 to: 1335

```

70 TCGCCGCTTACTTACCATATGCGTGTGCGGATTCGCCCAAGC 119
   ||| ||| ..... ||| ||| .....
44 SerGlyAlaValAlaGlyAlaLeuPheSerTyrProSerLeuAlaSerVa 60
120 CGGCGCGGACACACTTATTCGCGCATACCAATACATCTATGCGCACT 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
60 L...ValGlyAsnThr.....LeuProTribelThrTyrTrgAspR 73
170 TTGCGCAAAATPAAGGCAAGTTTGCATCGCGGCAAGATATGAGGTT 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
73 heaLaGluAsnLysGlyAlaPheHisLaGlyAlaThrAsnIleProLeu 89
220 TACACAAAAAAGGAGCTGTGCGCAATCGATGACGAAAGCCCGCAT 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 TyrGlyAlaGlyAsnGly...AlaValGlyGlyAlaGlyLeuAspLysAlaProMe 105
270 GATTGATTTTCTGTGTATCGCGT...AACGCGTGGCGGCAATTTGGCGG 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 LMetAspHisSerValAlaAspGlnIleLeuGlyValAlaIleThrLeuIleS 122
317 GCGATCATATATTTGTGAGCGTGCACATACGCGGCGCTATACATGTT 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 ePrGlnIleTyrValAlaGlyValLysHisAsnGlySerTyrAsnThrVal 138
367 GATTTGGTGGCGAGGAGCAATCCGATGACGACCGCTTTCTTACCA 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 ArgPheGlyTyrAlaAspAspThr.....ThyTyrTr 149
417 AATTGTGAAAAGAAATATATTAAGACGAGACTAACGCGCATCTTATG 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 GLeuValAspArgAsnGlu.....HisT 157
467 GCGGCGATATATCATATGCCGCTTGCAACAATTTGCACAGATGCAAA 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 rParGAspPheHisThrProAlaGlyLeuAsnLysLeuValThrGluValAla 173
517 CCTGTGAGATGACCAAGTATATGATGAGG...TGAATAACGCGTATTT 563
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174 ProValSerValThrAspAlaGlyThrGlyLysGlyValTyrGlnAsnAr 190
564 AAATAAATACCTGATCGTTGCAATGCGACGACGACGACACAAATTTG 613
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 gSerAlaGlyTyrProValPheTyrAlaGlyMetCylSerGlyThrGlnTyrTrng 207
614 GGTGCGATGACGACGACCAACCCATTAACCGGAAAGTTCAATCATATTTG 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
207 LysAlaIleSerGly.....AlaLeuThrArgIleLeu 217
664 ACGCATATTTCTGTGCTGCTGCGTGGCAATATCTTTGCACAAATGATG 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 GlyAlaTyrAlaTyrLysThrGlyGlyThrValGlySerProLeuIleSe 234
714 AGGTGCTGCACAGTCACACTTAGTAGCAAAAAATTAACATACCCCAT 763
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 rAspTrpSerLeuValSer...AsnProGlyTyrLeuTyrGlnSerVala 250
764 ATGGCTTTTACCAACAGAGGCTCATTTGGGACAGAGGCTGACACCATG 813
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 snGlyProLeuAlaSerTyrGlyThrProGlyAspSerGlySerProLeu 266
814 TTTATCATATGATGCCAAAGCAAAAGTGTATTAATATGAGGATTTGCA 863
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 PheAlaTyrAspAlaValLysLysGlnTyrValLeuValAlaValLeu... 282
864 AACAGCAACCCCTATATAGCAAAAGCAATAGCTTCCACGATGTTGCTGA 913
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

283 .....AsnGlyTyrAlaGlyGluLysGly.....LysT 292
914 AAGATTGCTCTATGATGAATCTTGTGAGATACCATTCAGTATTC 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
292 hrasnTrpPhe...ThrValIleProAlaGlyAspValAsnAsnThrIle 307
964 TACGAACACATCA.....AATGGAATA 989
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 LysGlnAspSerSerGlyThrValValProAlaValAlaGlyGlyLysPrI 324
990 CTTTTTTTAAAC...GACATATATATGCGCGAGCAAAATCGATGCCAAAC 1036
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 eValTrpAsnTyrSerGlySerGlyGlyGlnGlyThrLeuSerGlnAspG 341
1037 ATAACACATTTCTTACCTTATAGATTTAAACACGACACCGTTCAATG 1086
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
341 LysValIleTrpLysMet..... 346
1087 TTTAATGTTTCTTATTCGAGACAGACGAAACCTGTTATCATGCTGC 1136
346 ..... 346
1137 AGTGGGGTACACATTTATGCA...CCGACATGAAATATGCAAAATA 1183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 .....AsnGlyPheArgGlyGlySerLeuAsnAspGlyLysAspI 360
1184 TTTCCCTTATTTGACAAAGGAAAGTGATTTGATCTTACCACCAATC 1233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 LeuTrpPhe.....GlyGlyLysGlyTyrValValLeuLysAspAspVal 374
1234 AACCAAGCGCGGCGGCTTTGATTTTGTAGGATTAATTTTACGCTCGCC 1283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375 ValGlnGlyAlaGlySerLeuThrPheAsnGlyAspTyrThrValArgPr 391
1284 TAAACACACGAAACGTTGCAAGCGCGGCGCTTATATGATGATGCA 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 OGlnGlyAsnGlnThrTrpValGlyGlyGlyIleLeuAlaAsnAspLys 408
1334 GTACCGTTACTGGAAGTAAACGCGGTGCAACGACCGCTGTCCAA 1383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 LsArgValAspTrpMetValAsnGlyLeuAlaGlyAspAlaLeuHisLys 424
1384 ATCGCAAGGACAGCGTGTGTTCAAGCCAAAGGGAACCAAGGCTC 1433
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
425 ThrGlyLysGlyThrLeuValAlaIleGlySerGlyLysAsnProGlyTh 441
1434 GGTACGCGTGGCGGCGAGCTTAAGTCATCTTACATACACAGCGGACATC 1483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 rLeuAsnThrCylAspLysPheLysThrValIleLeuAlaGlnLysAlaAspAla 458
1484 AAGCAAAAAACAGCCTTAGTGAATCGCTGTGTCAGCGGACAGGCGG 1533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
458 lalGlyArgValArgAlaPheSerGlyValArgIleValSerGlyArgPro 474
1534 ACGGTGCACTGAATGCCGATATCACTGCAACCCGACAAACTATTT 1583
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
475 ValValIleLeuGlnAspSerHisGlnIleGluCylAspArgIleArgTr 491
1584 CGGCTTGGCGGCGGAGCTTGTGAATCGGCGGCTGCTTCTGCTTCC 1633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
491 rGlyTyrArgGlyGlyThrLeuAspIleAsnGlyAsnAspMetThrPhe 508
1634 ACCGATTTCAAAATACCGATGCAAGGCGGATGATTCACACCAATCAA 1683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
508 lAspGlyLeuAlaIleAlaAspLysGlyValAlaValLeuThrSer..... 521
1684 GACAAAGATTCACCGTTTACCATTTACAGGCAATPAAGATTTACTACAC 1733
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
522 ...ArgAlaGlySerAlaThrVal.....ArgLeuAspPheSerProse 535
1734 CGGCAATATACACAACTTGATACGCAAAAGAAATGCGTACACAGGTT 1783
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
535 rGly.....GlnLysAlaValMetTrpHisGly 545

```



```

|||||.....:|
997 SerProAlaProAlaIaIaSerValThrAspValProThrGlyGluAsnTh 1013
3525 ACCGACACCCCAACCGCAGCCGAC.....CTGATCAGCCGT 3562
1013 rGlyuSerGlyuTyArgAspGluThrArgTrpLeuThrGlyT 1030
3563 ATGCCAATAGCGGTTGAGTGAATTTCCGACGCTCAACAGGTTTTC 3612
1030 yr..... 1030
3613 GCCGTACAGACGAATTGACCGCGTGTTCGGAAGACCGCCGACGAC 3662
1031 .....ArgSerTh 1033
3663 CGTTTGACACAGCGCATCCGGACACCAACACTACCGTTCGCAAGTT 3712
1033 rValAsnIaSerAlaValaIArgAspIaGlyMetLeuMetSerMetGlyH 1050
3713 TCAGCGCCCTAC.....CGCCACAAACCGAC 3738
1050 lArGAsnPhelIleAsnGluValaAsnAsnLeuAsnIlyArgMetGlyAsp 1066
3739 CTGCGCCCAATCGTATGCGAGAAACCTC.....GG 3770
1067 LeuArGAspIleAsnGlyGluIaGlyIaTrpAlaArgIleMetSerG 1083
3771 CAGCGCGCGCGTGCATCCTTTTCGACACACCGGACCGGAAACACT 3820
1083 yThrGlySerAlaGlyGlyGlyPheSerAspAsnTrpThrAsnValGlnv 1100
3821 TCAGAC.....GACGCGATCGGCACACTCGGCACG 3849
1100 aIGlyAlaIaAspIlyuShisGluLeuAspGlyLeuAspLeuPheThrGly 1116
3850 CTTCGCCACGTCGCGTTTCGGCAATACGCGATCGCGAGTTGAC.. 3897
1117 ValThrMetThrTyThrAspSerAsnIaGlySerAspThrPheSerG 1133
3898 .....ATCGCATCAGCGCGCGCGCGGTTTACTA 3928
1133 yLysThrIySerValGlyAlaGlyLeuIyIaSerAlaLeuPheAspS 1150
3929 GCGGACGCTTTCAGACGCGATCAGAGCAAAATCCGCCGCGCTGCTG 3978
1150 erGlyAlaTrpIleAspLeuIle...GlyLys.....TyrVal 1161
3979 CATTCAGCATTCAGGCAAGATACGCGAGTTTCGGCGGATTCGGCAT 4028
1162 HlshIs.....AspAsnGluTyThrAlaThrPheAlaGlyLeuGlyTh 1176
4029 CGAAGCGCATTCGGCGCAACGCGATTTGTCGCAAAAGGATTTACC 4078
1176 rIyAspTrpIySerSerHisSerTrpTrpAlaGlyAlaGluValaGlyTy 1193
4079 GATAC.....GAAAACGTC 4092
1193 rGlyThrIaIaValThrGluAspAlaTrpIleGluProGlnIaGluLeuVal 1209
4093 AATATCGCAACCCCGGCTTCGATTCACACGCTACCGCGCGGCAATTAA 4142
1210 TyrGlyAlaValaSerGlyLysGlnPheSerTrpLysAspGlnGly.... 1224
4143 GGCAGATTTTCATTCAAACCGCGCAACACATTTCCATCAGCCCTTAT 4192
1225 .....M 1225
4193 TGAGCGCTCTATACCGATCGCGCTTCGGCAAAAGTCCGAACGCGCTC 4242
1225 etSerLeuSerMetLysAspIyAspTrpAsnProLeuIleGlyArgThr 1241
4243 AATACCGCGGTATGCGCGCATTTTCGCAAAACCGCGCATGCG...GA 4289
.....

```

1242 GlyVal.....AspValGlyLysSerPheSerGlyLysAs 1253

4290 ATGGGCGGTAAACGCC 4305

1253 pTrpLysValThrAla 1258

seq_name: sp_bacteria:Q54165

seq_documentation_block:

ID Q54165 PRELIMINARY; PRT; 1366 AA.

AC Q54165;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE SEPA PROTEIN PRECURSOR.

GN SEPA.

OS Shigella flexneri.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Shigella.

OX NCBI_Taxid=623;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=M90T;

RA MEDLINE=96020667; PubMed-7476198;

RA Benjeloun-Touimi Z., Sansonetti P.J., Parsot C.;

RT "Sepa, the major extracellular protein of Shigella flexneri;

RT autonomous secretion and involvement in tissue invasion.";

RL Mol. Microbiol. 17:123-135(1995).

CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

DR EMBL; 248219; CAA8252.1; -.

DR MEROPS; S06.0PW; -.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR000710; Iga_S6.

DR InterPro: IPR001254; Trypsin.

DR Pfam; PF02395; IGAL; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PROSITE; PS0240; TRYPSIN_DOM; 1.

KW Hydrolase; Serine protease; Signal.

FT SIGNAL 1 56 POTENTIAL.

FT CHAIN 57 1089 MATURE SEPA.

SEQUENCE 1366 AA; 146379 MW; 85EBB05F3E8A7D8 CRC64;

alignment_scores:

Quality: 996.00 Length: 1599

Ratio: 1.233 Gaps: 64

Percent Similarity: 50.532 Percent Identity: 25.203

alignment_block:

us-09-303-518d-653 x Q54165 ..

Align seg 1/1 to: Q54165 from: 1 to: 1366

```

22 ACAACCGAACAACACCGCAAAAGCCCTAAACCGCGCGCATCGCTTCTC 71
||| .....:|
27 ThrGlySerHisArgArgLeuSerArgArg...ValIleuThrSe 42
72 GCCCGCTTACTAGCATATGCTGCTGCTGCTGCGCATTCGCCCAACCC 121
||| .....:|
42 rValAlaIaLeuSerLeuSerSerAlaTrpProIaIaSerAlaThrV 59
122 GGGGCGGACACACTTATTCGATCACTACCAATACATACGCGACTT 171
.....:|
59 alSerAlaGlu.....IleProTyrglnIlePheAlaGAspPhe 71
172 GCCGAAATATAAGCAAGTTTGACGTGCGGCGGAAGATATTGAGTTTA 221
|||||.....:|
72 AlaGluAsnLysGlyGlnPheThrProGlyThrThrAsnIleSerIleTy 88
222 CAACAAAAAGGAGGTTGCTCGCGCAATATGATACGAAACCCCGCATGA 271
|||||.....:|
88 rAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysAlaProMet 104
.....

```

```
272 TTGATTTTCTGTGATCGCTAACGGCGGTGGCGCATTTGGCGGCGAT 321
|||||
104 LaspheSerSerAlaThrIleThrThylGlySerLeuProProGlyAsp 120
322 .....CAATATATTGTAGCGGTGGCGCATTAACGGCGG 353
|||||
121 HisThrLeuYrSerProGlnYrValValThrAlaLysHISValSerGI 137
354 CTATACCAATGTGATTTGTGGCGGAGGAGCAATCCCGATCAGCACC 403
|
137 YSerAspThrMetSerPheGly..... 144
404 GCTTTTCTTACCAATTTGAAAAGCAATATATTAAGCA...GGGACT 450
|||||
145 .....TyrAlaLysAsnThrTyrThrAlaValGlyThr 155
451 AACGGCCATCTTATGCGCGCATTAATCCGCGTTTGCACAAAT 500
|||||
156 AsnAsnAsnSer...GlyLeuAspIleYrThrArgArgLeuSerLysIle 171
501 TGTACAGATGCGAAGCCTGTGAGATGACC..... 531
|||||
171 uValThrGluValAlaProAlaGluValSerAspIleGlyAlaValSerG 188
532 ..AGTTATATGATGGGTGAATACGCTGATTTAAATTAATACCTGAT 579
|||||
188 lYAlaYrGlnAlaGlyArgPheThrGluPhe..... 199
580 CGTGTTCGATCGAGCAGCAGCAGCAATATATGGCGCTGATGAGACGA 629
|||||
200 ...TyrArgLeuGlyGlyMetGlnValYrValYs.....AspIly 212
630 ACCCAATTAACCGGAAAGTTCAATCATATTCAGCGCATTTCTGGC 679
|
212 sAsnGlyAsnArgThrGlnValYr.....ThrAsnGlyGlyPheI 226
680 TGTGTGGGTGGCAATACCTTGCACAAATGATCAGGTGGTGGC..... 723
|||||
226 euValGlyGlyThrValSerAlaLeuAsnSerTyrAsnAsnGlyGlnMet 242
724 ...ACAGTCACTAGTACGCAAAAATTAACATAGCCCATATGCTT 770
|||||
243 lIethrAlaGlnThrGly...AspIlePheLeuIleProProAsnGlyPyr 258
771 TTTACCAACAGGAGGCTCATTTGGCAGACAGTGCACCAATGTTATCT 820
|||||
258 oleuAlaAsnYrLeuAsnMetGlyAspSerGlySerProLeuPheAlaI 275
821 ATGATGCCCAAAAGCAAAAGTGTAAATTAATGGGTATTCGAAACAGGC 870
|||||
275 YrAspSerLeuGlnYrLysIleValLeuIleGlyValLeuSerSerGly 291
871 AACCCCTATATGAGAAAAGCAATGCTCCACAGTTCGTAAGATG 920
|
292 ThrAsnYr.....GlyAsnAsnTyrValValThrGlnAspPh 305
921 GTTCTATGATGAATCTTGTCTGAGATACCATTCAGTATTTCTACGAAC 970
|
305 eleu..... 306
971 CACATCAAAATGGGAATACTTTTAAAGCAATATTAATGGCGCAGGA 1020
|||||
307 .....GlyGlnGlnProGlnAsnAsp..... 313
1021 AAATGATGCGCAACATTAACACATATTCCTACCTTATAGATTAATAAC 1070
313 ..... 313
1071 ACGAAGCTTCAATGTTTAAATGTTTCTTATTCGAGACAGACAGACAG 1120
|||||
314 .....PheAspLysThrIleAlaYrThrSerGlyGlnG 325
1121 CTGTT.....TATCATGCTGAGGTGGCGCTCAAC..... 1149
|||||
325 lYValLeuGlnTrpLysTyrAspAlaAlaAsnGlyThrGlyThrLeuThr 341
1150 .....AGTTATGACCCAGCA.....CT 1166
|
342 GlnGlyAsnThrThrTrpGlyTyrAlaIleTrpLysGlnArgLysLeuIle 358
1167 GAATTAATGGAAAAATATTCTTATTGACAAAGAAAGCAATGTA 1216
|
358 uAsnAlaGlyLysAsnLeuLeuPheThrGly...AsnAsnGlyGlyValY 374
1217 TACTTACAGACATCAACCAAGCGCGCGGTGTGATTTGAGGCT 1266
|
374 AlLeuGlnAsnSerValAsnGlnGlyAlaGlyYrLeuGlnPheAlaGly 390
1267 AATTTACGGTCTCGCTAAACAAACAGCAAGCGCGCGGCGCT 1316
|
391 AspTyrArgValSerAlaLeuAsnGlyGlnThrTrpMetGlyGlyGlyI 407
1317 TCATATCAGTATGCGAGTACCTTACTTGGAAAGTAAACGGGTGCAA 1366
|
407 eIleThrAspLysGlyThrHisValLeuTrpGlnValAsnGlyValAlaG 424
1367 ACGACCGCTGTCCAAATGCGCAAGCAGCGCTGTTCACGCCAAA 1416
|
424 lYAspAsnLeuHisLysThrGlyGlnGlyThrLeuThrValAsnGlyThr 440
1417 GGGGAAACCAAGGCTGCTGACAGGTGGCGGCGAGCGTAAGCTATCTAGA 1466
|
441 GlyValAsnAlaGlyGlyLeuLysValGlyAspGlyThrValIleLeuAs 457
1467 TCACAGCGGAGCATCAAGCAAAAACAGCCTTATGGAATGGCT 1516
|
457 nGlnGlnAlaAspArgAspGlyLysValGlnAlaPheSerSerValGlyI 474
1517 TGTCTACGCGCAGGCGGAGCTGCAACTGAATCCGATATATCGTTCAAC 1566
|
474 leAlaSerGlyArgProThrValValLeuSerSerPheGlnValAsn 490
1567 CCCGCAACAACCTATTTTGGCTTTCGGCGGAGCGTGTGATTTGAACGG 1616
|
491 ProAspAsnIleSerTrpGlyTyrArgGlyYrArgGlnLeuAsnGln 507
1617 GCATTCGCTTTCGTCACCGCATTCACAAATACCGATGGAAGCGCATGA 1666
|
507 YrAsnAsnLeuThrPheThrArgLeuGlnAlaAlaAspTyrGlyAlaIle 524
1667 TGTCAACCACAATCAAGCAAAAGAAATCCACGCTTACCAAT 1707
|
524 lethrAsn...AsnSerGlyLysSerThrValThrLeuAspLeuGln 539
1708 ...ACAGGCAATTAAGATATTACTACACCGGCAATTAACACACTTG 1752
|
540 ThrLeuLysAlaSerAspIleAsnValProValAsnThrValSerIlePh 556
1753 .....GATAGCAAA..... 1761
556 eGlyGlyArgGlyAlaProGlyAspLeuYrTyrAspSerSerThrLysG 573
1762 .....AAGAATTTGCTACACGCTTGTGGTGGCGAG... 1794
|
573 lntyrPheIleLeuLysAlaSerSerTyrSerProPheSerAspLeu 589
1795 .....AAGATGCAACCAAAAG... 1812
590 AsnAsnSerSerValTrpGlnAsnValGlyLysAspHisAsnLysAlaI 606
1812 ..... 1812
606 eAspThrValLysGlnGlnLysIleGluAlaSerSerGlnProTyrMet 623
1813 .....AACGGCGGCTCAATCTGAATTAACCAACCGGAGAA 1848
```

```

623 yrhlsGlyInLeuasnGlyasnMetAspValasnlleProGlnLeuSer 639
      |||||.....:
1849 GCGGATCCACATTACTCTTCCTCCGCGGAAACAATTA..AACGGCA 1855
      :||| :|||:|||||:
640 GlyAspValleuAlaLeuaspPylSerValasnlleProGlnGlySe 656
      :|||:|||||:
1896 TATCAGCGCAACAACGCGCAATGTTTTCAGCGGAGACCGACCGC 1945
      :|||:|||||:
656 rlethrlyslserserGlythrleuIlePheGlnIleHisProValIleH 673
      :|||:|||||:
1946 ACGCCTACACATTATTAGCAAGCGGTGTCAAAAATGGAAGATGCCA 1995
      :|||:|||||:
673 lsalaglyThrthrThrserSerGlnSer..... 683
1996 CAAGGAGAAATCGTGTGGACAGATTGATCCAGCCGACATTAAAGC 2045
      |||||
684 .....AspTrpGlnThrArgGlnPheThrLe 692
2046 GGAATACTCATATTTCAGGCGGACAGCGGTGTTCCGCAATGTTG 2095
      |||:|||||:
692 uGlnlyslleuylsleuaspAlaIaIaThrPheHisleuSerArgash..G 708
2096 CCNAAGTGAAGGCGATTGGCATTTAAGCAATCAGCCCAAGCAGTTTTC 2145
      :|||:|||||:
708 lylsMetGlnGlyaspIleasnIaIaThrasnGlySerThrValIleleu 724
2146 GGTGTGCAACCGCATCAAGCCACACATCTGTACAGCTTGGAGTGGAC 2195
      :|||:|||||:
725 Gly.....SerSerArgValPheThrAspArgSerAspGlyTh 737
2196 GGGTGTGACA.....AGTTGTACGGAATAAACCATTA 2227
      |||:|||||:
737 rGlysnAlaValSerSerValGlnGlySerAlaIaIaThrAlaIaThrValG 754
2228 CCGACGATTAAGTGAATTCCTTCATTAGACAGACCGACATCAGACGCAAT 2277
      :|||:|||||:
754 LyAsp.....GlnSerAspTyrSerGlyasn 762
2278 GTACACCTTGGCGATCAGCTCATTTAATCTCACAGACCTTGCACACT 2327
      :|||:|||||:
763 ValThrleuGlnAsnLysSerSerleuGlnIle.....MetGlnArgPhe 777
2328 CAACGCGCAATCTTAGTCAGCGGACGACGACATATAGGTTAGCGCCA 2377
      :|||:|||||:
777 ethrlelyGlyIleGlnAlaIaTyrAspSerThrValSerValThrSerGlnA 794
2378 ACGCCAC.....CAAAACGCGCAACCTCAGCCTC 2406
      :|||:|||||:
794 snAlaValPheAspArgValGlySerPheValasnlSerSerleuThrleu 810
2407 GTGGGCAATGCCCAAGCAACATTTAATCAGCCACATTAAC..GGCAA 2453
      :|||:|||||:
811 GlyLysGlyAlaLysleuThrAlaGlnSerGlyIlePheSerThrGlyAl 827
2454 CACATCGGTTGCGACATGCTTCATTATTAATTAAGCAAC.....AAGC 2497
      :|||:|||||:
827 aValaspValLysGlnAsnAlaSerleuThrleuThrleuIleMetProSerA 844
2498 CCGTCAAAAACGCG.....AGCTGACGCTTTCGACAAACGCTAAGCA 2541
      :|||:|||||:
844 lAGlnLysGlnGlyTyrTyrSerProValIleSerThrThrGlnGlyLe 860
2542 AACGTAAAGCATTCGCGACCTCAACGCGCAATGTCTCCCTAAGCGAATAAGC 2591
      :|||:|||||:
861 AsnleuGln..... 863
2592 AGTATTCATTTTGAACACGCGCTTTTACCGGAAAAATCAGCGCGCA 2641
      :|||:|||||:
864 .....AspAsnAlaSerPheSerValLysAsnMetGlyTyrL 876
2642 AGGATACGGCATTTACCTTAAAGACAGCGCAATGACGCTGCCGTGGGC 2691
      :|||:|||||:
876 euSerSerAspIleHis.....AlaGly 883
2692 ACGGAATTGGCAATTTAAACCTTGCACAGCCGACCATTTACACTCAATTC 2741
      :|||:|||||:
884 ThrThrAlaIaIaThrIleasnleuGlyAspSer..... 894
2742 CGCCTATTCGACACGATGCGCGACGCGCAAAACCGGACGTGGCGAGATG 2791
      :|||:|||||:
895 .....AspAlaAspAlaGlyLysThrAspSerProleuPheS 907
2792 CGCGCGCGCGCGCTTCGCGCGCTTCCTCATTA.....TCCGTTAGCGG 2835
      :|||:|||||:
907 erSerleuMetLysGlyTyrAsnAlaValleuArgIleSerIleThrGly 923
2836 CCAACTTCGCGCAGATCCGCTTCACACACGCTGACGCTAATC..... 2877
      :|||:|||||:
924 AlaGlnSerThrValasnMetIleAsnAlaIaIaLeuTrpTyrSerAspLys 940
2878 .....GGCAATTTGAACGGTCAAGGACATTCGCTTATGTGCG 2917
      :|||:|||||:
940 sSerGlnAlaGlyAlaIaIaIaLysGlySerArgIleGlnleuGlyA 957
2918 AACTCTTCGCGCTACCGCGCGCAATTTGAAGTGGCGGAAATTCGCA 2967
      :|||:|||||:
957 spGlyLysHisPheAlaIaIaThrleuGlnValLysGlnleuSerAlaAspAsn 973
2968 GGCACCTACACCTTGCTGTCACAAATACCGCGCAACGCAACCCGTAAGTCT 3017
      :|||:|||||:
974 ThrThrThrleuMetHisThrAsnAsnSer.....ArgAl 985
3018 CGACCAATTTGACGCTAGTGTGAA.....GGAATAAGACACACACCGC 3058
      :|||:|||||:
985 aAspGlnleuAsnValThrAspLysleuSerGlySerAsnAsnSerValI 1002
3059 TGTCGGAATCTTATTTTCAACCTGCAAAACGACACGCTGCATGCCGCG 3108
      :|||:|||||:
1002 euValAspPheleuAsn..... 1007
3109 GCATGGCGTTATACGTTATCCGCAAGACGCGAGTCCGCTGCATAA 3158
      :|||:|||||:
1008 .....Ly 1008
3159 TCCGCTCAAAAGACAGAGCTTTCGCAAACTCGGCAAGCGGAGGAAA 3208
      :|||:|||||:
1008 sProAlaSerGlnMetSerValThr..... 1016
3209 CAGAGCGCGCTTGACGCGCAAAACGACACACTTGCAGCAACACAG 3258
      :|||:|||||:
1016 ..... 1016
3259 GCGGAAAAAGACAGCGCAAGCGCTTGACGCGCTGATTCGCGCGGCGC 3308
      :|||:|||||:
1017 .....leuIleThrAlaProLy 1022
3309 CAATGCCACGCAAAAGCAGAAAGTGTGCCGAACCGCGCGCAGCAG 3358
      :|||:|||||:
1022 sGlySerAspLysThrPheThr.....AlaG 1032
3359 GCGGGAATAATGCGCGCTTATGACGCGGAGGAGAAAGAAAAAGCGGTG 3408
      :|||:|||||:
1032 lYthrGlnGlnIleGlyPheSerAsnValThrProValIleSerThrGlu 1048
3409 CAGCGGATTAAGACAGCGCGCTTGCGAAACAGCGCAACGGAACCCG 3458
      :|||:|||||:
1049 LysThrAspAspAlaThr.....LysTrpValleuThrGln 1060
3459 GCGGCTACACGCGCTTCCCGCGCGCGCGCGCGCGCGGATTTGC 3508
      :|||:|||||:
1060 yTyrGlnThrThrAlaAspAlaGlyAlaSerLysAlaIaLysAsp.... 1075
3509 CGCAACGCGCACCCCAACCGCAACCCCAACCGCGCGCGCATGATCAGC 3558
      :|||:|||||:
1075 ..... 1075

```

```

3559 CGTTATGCCATAGCGGCTTTGAGTGAATTTCCGCCACGCTCAACAGCGT 3608
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1076 ...PhemetalaserGlyTyrLysSerPheLeuThrGluValAsnAsnLe 1091
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3609 TTTCGCCGTCACAGCAGCATTTGACCGCGCTTTTGGCCGAACCGCGCGA 3658
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1091 uasnLysArgMetGlyAspLeu.....ArgAspThrGlnG 1103
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3659 AC.....GCCGTTTGACAAAGCGGATCGGCGACACAAACATACGCT 3702
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1103 LyspAlaGlyValThrAlaArgLleMetAsnGlyThr.....Gly 1116
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3703 TGGCAGAGATTTCCGCCCTACCGCCACAAACGACTTGGCCAAATCGG 3752
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1117 SerAlaAsp...GlyAspTyrSerAspAsnTyrThrHisValGlnIleG 1132
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3753 TATCGAGAA.....AACCTCGCGACGCGGCGCG 3781
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1132 yAlaAspArgLysHisGlyLeuAspGlyValAspLeuPheThrGlyAlaL 1149
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3782 TGGCAGATCGCTGTTTGGCAACACCGGACCGGAACACCTTC..... 3822
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1149 euLeuThrTyrThrAspSerAsnAlaSerSerHisAlaPheSerGlyLys 1165
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3823 GACGACGCGATCGGCACTCGGCGAGCGCTTGGCCAGCGCTTTTGGG 3872
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1166 AsnLysSerValGlyGly..... 1171
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3873 GCAATACGCGATCGGACAGTTTGCACATCGCATCGACGCGGCGCGGTT 3922
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1172 .....GlyLeuThrAlaSerAlaLeuP 1179
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3923 TTAGTAGCGGCGACGCTTTCACAGCGCATCAGGCAAAATCCGCGCGC 3972
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1179 heAsnSerGlyAlaTyrPheAspLeuThr.....GlyLys..... 1190
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3973 GTGCTGCATTAAGCGCATTCAGGCAAGATACCGCGAGTTTCGCGGAT 4022
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1191 TyrLeuHisHisAspAsnGln.....HisThrAlaAsnPheAlaSerLe 1205
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4023 CGGCTGCACACCGGCATCGCGCGACCGGCTATTTGCTCCAAAAGCGG 4072
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1205 uGlyThrTyrAspTyrSerSerHisSerThrTyrAlaGlyAlaGluValG 1222
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4073 ATTACCGATAC.....GAA 4086
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1222 LyrTyrAlaTyrHisLeuThrLysGluSerTyrValGluProGlnIleGlu 1238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4087 AACGTCATATATCGCACCGCGCTTGCATTAACCGCTACCGCGCGG 4136
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1239 LeuValTyrGlySerValSerGlyLysAlaPheSerThrGlyAlaArgL 1255
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4137 CATTAAGCGACATTAATTCATCAAAACCGCGCAACACATTTCCATCAGC 4186
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1255 yMetAla.....LeuSerMetLysAspLysAspTyrAsnProLeuIleG 1270
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4187 CTATATGAGCGCTGCTCTATACGCATCGCGCTTCCGCG.....AAA 4227
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1270 LyrAlaGlnGlyValAspValGlyAlaGlyAlaPheSerThrGlyAlaArgL 1286
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4228 GTCCGAACGCGC.....GTCAATAC 4247
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1287 IleThrAlaThrGlyAlaGlyLeuGlyTyrGlnPheAspLeuLeuAlaAsnG 1303
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4248 CGCGGTTTGGCGAGATTTCCGCAAAACCGCACTGCGGAATGG.... 4293
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1303 yGluThrValLeuGlnAspAlaSerGlyGlyLysArgPheGluGlyLys 1320
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4294 .....GCCGTAACGCGCGCAATCAAA 4314
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1320 yAspSerArgMetLeuMetThrValGlyMetAsnAlaGluIleLys 1335

```

```

seq_name: sp_bacteria:099QC6
seq_documentation_block:
ID 099QC6 PRELIMINARY; PRT; 1364 AA.
AC 099QC6;
DT 01-JUN-2001 (Tremblrel, 17, Created)
DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE SEPA. EXTRACELLULAR SERINE PROTEASE OF THE IGAL PROTEASE FAMILY,
DE SECRETED BY A C-TERMINAL AUTOTRANSFERTER DOMAIN (SECRETED
DE PROTEASE).
GN SEPA.
OS Shigella flexneri.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T;
RX MEDLINE=20566792; PubMed=1115111;
RA Buchrieser C., Glaser P., Rusniok C., Nedjari H., d'Hauteville H.,
RA Kunz F., Sansonetti P., Parsot C.,
RT "The virulence plasmid pWR100 and the repertoire of proteins secreted
RT by the type III secretion apparatus of Shigella flexneri.";
RL Mol. Microbiol. 38:760-771(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID-VIRULENCE pWR501;
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,
RA Blattner F.R.;
RT "Complete DNA Sequence and Analysis of the Large Virulence Plasmid of
RT Shigella flexneri.";
RL Infect. Immun. 0:0-0(2001).
DR EMBL; AL391753; CAC05786.1;
DR EMBL; AF348706; AAK18385.1;
DR MEROPS; S06;
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF02395; IGAL1.
DR PRINTS; PR00921; IGASERPTASE.
DR SMART; SM00020; TRYP_SPE; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
KW Hydrolase; Plasmid; Protease; Serine protease.
SQ SEQUENCE 1364 AA; 145937 MW; 71165A82BD09B9E CRC64;

```

```

alignment_scores:
Quality: 995.00 Length: 1587
Ratio: 1.234 Gaps: 62
Percent Similarity: 50.788 Percent Identity: 25.079

```

```

alignment_block:
US-09-303-518D-653 x 099QC6 ..

```

```

Align seg 1/1 to: 099QC6 from: 1 to: 1364

```

```

22 ACACGGAACACACCGCAAGCCCTTAACCGCGCGATCGCTTC 71
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27 ThrCysLysSerHisArgTyrGlySerLysArgTyr...ValIleLeuThrSe 42
72 GCCCGTTACTAGCCATATGCTGTCGTTCGGCATTTCTGCCCAAGCC 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42 rValAlaIleLeuSerLeuSerSerAlaTyrProAlaLeuSerAlaThrV 59
122 GCGCGGACACACACTTATTTGGCATCACTACCAATCTATTCGCGACTT 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 alSerAlaGlu.....IleProTyrGlnIlePheThrAspPhe 71
172 GCCGAAATTAAGCGCAAGTTTGCACGCGCGCGCAAGATATTAGGTTTA 221
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 AlaGluAsnLysGlyLysPheThrProGlyThrThrAsnIleSerIleTyr 88
222 CAACAAAAAAGGCGAGTTGTCGGCAATCGATGACGAAAGCCCGCATGA 271

```

```

1103 CCGAGACAGCAAGAACCTGTTATCATCTGTCAGGTGGGTCACAGT 1152
349 .....HisGlyIysIysGlyAsnAsp... 355
1153 TATGACCCAGACTGAATTAATGGAGAAATATTCCTTTATGACAAAG 1202
356 .....LeuAsnAlaGlyIysAsnLeuLeuPheThrGly...As 367
1203 AAAAGGTGATTTGATCTTACACAGCACATCAACCAAGCGCGCGCGT 1252
367 nasnglygluvalvalleuInAsnSerValasngInglyalaglytyrl
1253 TGTAFTTTGAGGGTAATTTTACGGTCTCGGCTAAACACGAAACGTG 1302
384 eugInPheAlaGlyAspItyrArgValSerAlaLeuAsnnglyInThrTTP
1303 CAAGCGCGGGCGTTTCATATCATGATGACAGTACCGTCTACTTGAAGT 1352
401 MetGlygluIleIleThrAspIysGlyThrHisValLeuTrpGlnVala
1353 AACGGCGGTGGCAACACGCGCTGTCCAAATCGCAAAAGCACGCTGC 1402
417 lasnglyvalalaglyAspAsnLeuHisIysThrGlygluGlyThrLeuT
1403 TGGTTCAAGCCAAAGGGGAAACCAAGGCTCGGTCAGCGTGGGAGCGT 1452
434 hrValasnglyThrGlyValAsnAlaGlyIleuIysValGlyAspGly
1453 AAAGTCATCTTAGATCAGCAGCGAGCATCAAGGCAAAAACAGCCTT 1502
451 ThrValIleLeuAsnngInGlnAlaAspAlaAspIlyValGlnAlaPh
1503 TAGTGAATTCGGCTTGGTCAGCGGAGGGGAGCGTGCACACTGAATCCG 1552
467 eSerSerValGlyIleAlaSerGlyArgProThrValValLeuSerAsps
1553 AFAATCAGTTCAACCCGACCAACCTATTTGCGCTTGGCGGGAGCGT 1602
484 erGlnGlnValAsnProAspAsnIleSerTrpGlyTyrArgGlyGlyArg
1603 TTTGATTTGAACGGGCGATTCGCTTTCGTCACCGCATTCAAAATACGA 1652
501 LeuGlnLeuAsnnglyAsnAsnLeuThrPheThrArgLeuGlnAlaAs
1653 TGAAGGGCGATGATGTGTCAACCAATCAAGAAAGAAATCCACCTTA 1702
517 pTyrlValAlaIleIleThrAsn...AsnSerGlnIlySerThrValT
1703 CCAT.....ACAGCAATAAGATATTACTACACCGGCAT 1740
533 hrLeuAspLeuGlnThrLeuLysAlaSerAspIleAsnValProValAsn
1741 AACACAACTTG.....GA 1754
550 ThrValSerIlePheGlyArgGlyAlaProGlyAspLeuTyrTyrAs
1755 TAGCAAA.....AAGAATTCGCTACCAACGGTT 1783
566 pSerSerThrLysGlnTyrPheIleLeuLysAlaSerSerTrpProp
1784 GATTTCGCGAG.....AAGAT 1800
583 hePheSerAspLeuAsnAsnSerSerValTrpGlnAsnValGlyLysAsp
1801 GCACCAAAACG..... 1812
600 HisAsnLysAlaIleAspThrValLysGlnInLysIleGlnAlaSerSe
1813 .....AAGGGCGGCTCAATCTGATT 1834
616 rGlnProTyrMetIyrlHisGlyGlnLeuAsnnglyAsnMetAspValAsnI

```

```

1835 ACCAACGGAGAGCGGATTCGACTTACGCTTCCGGCGACAAAT 1884
      ::::: ::::: ::::: :::::
633 LeproglInleuSerGlyAspValLeuAlaLeuAspLysSerValAsn 649
1885 TTA...AACGGCAATATCACGCAACAAACGCAACTTTTTCAGCGG 1931
      ::::: ::::: ::::: :::::
650 LeproglInGlySerIleThrLysLysSerGlyThrLeuIlePheGlnI 666
1932 CAGACGGACACGGCAGCTTACATCATTTAGGAGCGGGTGTGCAAAA 1981
      ::::: ::::: ::::: :::::
666 yHISProValIleHisAlaGlyThrThrSerSerSerGlnSer.... 681
1982 TGGAGGTATCCCAAGAGAAATGCTGTGGACAAAGATTGGATCGAC 2031
      ::::: :::::
682 .....AspTrpGluThr 685
2032 CGCACATTTAAAGCGGAAACTTCATATTCAGGGCGGCAAGCGGTGT 2081
      ::::: ::::: ::::: :::::
686 ArgGlnPheThrLeuGlnLysLeuAspAlaAlaThrPheHisLe 702
2082 TTCCCGCAATGTTGCCAAAGTGAAGCGATGGCATTTAGCAATACAG 2131
      ::::: ::::: ::::: :::::
702 uSerArgAsn...GlyLysMetClnGlyAspIleAsnAlaThrAsnGlyS 718
2132 CCCAACGACTTTTGGGTGTCGACGCGCATCAAGCCACAACTGTACA 2181
      ::::: ::::: ::::: :::::
718 eThrValIleLeuGly.....SerSerArgValPheThrAsp 730
2182 CGTTCGACTGACGGGTCTGACA.....AGTTGAC 2213
      ::::: ::::: :::::
731 ArgSerAspGlyThrClnAsnAlaValSerSerValGlnGlySerAla 747
2214 CCAAAAAACATTTACGACGATAAAGTATGCTTCATTGACGACAGCG 2263
      ::::: ::::: :::::
747 rAlaThrThrValGlyAsp.....GlnSerA 756
2264 ACATCAGAGCAATGTACGCTTGGCGATCAGCTCATTTAAATCTACA 2313
      ::::: ::::: :::::
756 sPLysSerGlnAsnValThrLeuGlnAsnLysSerSerLeuGlnIle... 771
2314 GGACTTGGCCACACTCAGCGCAATCTTAGTCAGGCGGAGACACCACTA 2363
      ::::: ::::: :::::
772 ...MetGluArgPheThrGlyIleGluAlaIlyAspSerThrValSe 787
2364 TACGGTTACGCGCAACGCCACC.....CAAACG 2392
      ::::: ::::: :::::
787 rValThrSerGlnAsnAlaValPheAspArgValGlySerPheValAsn 804
2393 GCAACCTCAGCCTCGTGGGCAATGCCCAAGCAACATTTAAATCAAGCA 2442
      ::::: ::::: :::::
804 eSerLeuThrLeuGlyLysGlyAlaIlyAspLeuThrAlaGlnSerGlyLe 820
2443 TTAAC...GGCAACACATCGGCTTCGACACATGCTTATTAATCTAAG 2489
      ::::: ::::: :::::
821 PheSerThrGlyAlaValAspAllyLysGlnAsnAlaSerLeuThrLeuTh 837
2490 CAAC.....AAGCGCTACAAAGCGC.....AGTCTGACGCTTTCG 2527
      ::::: ::::: :::::
837 rGlyMetProSerAlaGlnLysGlnGlyTyrTyrSerProValIleSerT 854
2528 ACAACGCTAAGGCAACGTAAAGCATTCGCACTCAACGGAATGTCTCC 2577
      ::::: :::::
854 hThrGlnGlyIleAsnLeuGlu..... 861
2578 CTAGCCGTAAGCGCAGTATTCATTTTAAACAGCCGCTTACCGGAAA 2627
      ::::: ::::: :::::
862 .....AspAsnAlaSerPheSerValIly 869
2628 AATCAGCGCGCGCAAGATACGCGATTACCTTAAAGACAGCGAATGGA 2677
      ::::: :::::
869 sAsnMetClnLysThrLeuSerSerAspIleHis..... 879

```

```

2678 CGCTGCCGCGGACGCAATTAGGCAATTTAACTTTGACAAACGCCACC 2727
      ::::: ::::: :::::
880 .....AlaGlyThrThrAlaAlaThrIleAsnLeuGlyAspSer... 892
2728 ATTACACTCAATTCGGCTATCGACAGATGCGGACGCGGCAACCGG 2777
      ::::: :::::
893 .....AspAlaSerAlaGlyLysThrAs 900
2778 CAGTGGCGCAATGCGCGCGCGCGCTGCGCGCTTCCCTATTA.... 2823
      ::::: :::::
900 pSerProLeuPheSerSerLeuMetLysGlyTyrAsnAlaValLeuArg 917
2824 ..TCCGTTACGCGCGCAACTTCGCGCAGAAATCCCTTCAACAGCGTACG 2871
      ::::: :::::
917 LysThrIleThrGlyAlaGlnSerThrValAsnMetIleAsnAlaLeuTrp 933
2872 GTAAAC.....GGCAAAATGAACGGTCAGGGAACATT 2903
      ::::: :::::
934 TyrSerAspGlyLysSerGlnAlaGlyAlaLeuLysAlaLysGlySerA 950
2904 CCGCTTATATGCGGAGACTTCGCGCTACCGGCGGCAAAATGAAGCTG 2953
      ::::: :::::
950 gIleGlnLeuGlyAspGlyLysHisPheAlaThrLeuGlnVallyGluL 967
2954 CGAAAGTTCCGAAGCGCACTTACACTTGGCTGTCAACAATACGCGCAC 3003
      ::::: :::::
967 eSerAlaAspAsnThrThrPheLeuMetHisThrAsnAsnSer..... 981
3004 GAACCCGTAAGCTTCGACCAATTGACGCTAGTGGA.....GAAA 3044
      ::::: :::::
982 .....ArgAlaAspGlnLeuAsnValThrAspLysLeuSerGlySe 995
3045 AGACACACACCGCTGTCCGCAAAATCTTAATTCACCCCTCAAAACGAAC 3094
      ::::: :::::
995 rAsnAsnSerValLeuValAspPheLeuAsn..... 1005
3095 ACGTCGATGCGCGGCGCATGGCGTTATCGCTTATCCGCAAGACGCGAG 3144
      ::::: :::::
1005 ..... 1005
3145 TTCGCGCTGATAATCCGCTCAAAAGAACAGACCTTCCGACAACTCGG 3194
      ::::: :::::
1006 .....LysProAlaSerGlnMetSerValThr..... 1014
3195 CAAGGCGCGAAGAAACAGAGCGCGCTTGACGGCAAAACGACAACTTG 3244
      ::::: :::::
1014 ..... 1014
3245 CCGCCAAACAAACAGCGCGAAGAAAGACACGCGCAAGCCTTGACGCGTG 3294
      ::::: :::::
1015 .....Leu 1015
3295 ATTGCGCGCGCGGCGCAATGCCACCGAAAGGCAAGAAAGTGTCCGCAAC 3344
      ::::: :::::
1016 IleThrAlaProLysGlySerAspGlnLysThrPheThr..... 1028
3345 GCGCGCGGACGAGCGGGAATGCGGCAATTAAGCGGCGGAGGAGANG 3394
      ::::: :::::
1029 .....AlaGlyThrGlnGlnIleGlyPheSerAsnValThrProy 1042
3395 AGAAAAAACGGGTGACGCGCGATTAAGACACCGCTTGGCGAAACAGCGC 3444
      ::::: :::::
1042 allIeSerThrGlnLysThrAspAspAlaThr.....Lys 1053
3445 GAAGCGGAACCCGCGCGGCTACACCGCTTCCCGCGCGCGCGCGC 3494
      ::::: :::::
1054 TrpValLeuThrGlyTyrGlnThrThrAlaAspAlaGlyAlaSerLysAl 1070
3495 CGCGCGGATTTGCGCAACCGCACGCGCAACCGCAACCCCAACCGCAC 3544
      ::::: :::::
1070 aAlaLysAsp..... 1073
3545 GCGACCTGATCAACCGCTTATGCCAATAGCGGTTTGAATGATTTTCGCC 3594

```



```

499 TTTGTCAGATGACAAACCTGTGATGACAGTATATGATGGTGC 548
      ::::::::::: |||:::
160 LeuValThrGluValIleProValAsnThrGly..... AspGly.. 172
549 GAAATACGCTGATTAAATAATACCCTGATGCTTGCATCGAGAG 598
      |||||:::
173 .LysTyrAsnAspSerLysLysThrAlaPheTyrArgThrGlyValG 189
599 GCAGACAATATTGGCGCTGTGATGAAGACCAATACCGCAAGT 648
      || |||||:::
189 LysAspIntyrIleLys..... AspLysLysGlyAsnHisThrLeu 202
649 TCATATCATATTGCAAGCGCATATTCTTGCTGCTGGTGC..... 690
      ::::
203 LeuAsnArgTyrAlaProGlnAsnAlaTyrLeuThrGlyThrValG 219
691 .....
219 YThrProTyrTyrGlyArgValSerAsnGlnThrValLeuIleSerSerP 236
716 GTGTGCGACAGTCACTTAGTACGAAAAAATAAACATAGCOCATAT 765
      |||:::| |||
236 roGlyAsnThrPheAsn..... GlnThrHis 244
766 GGTTTTATACCAACAGAGGCTCATTTGGCGACAGTGCATCCAAATGT 815
      |||:::|:::
245 GlyProLeuAlaSerTyrGlyGlnSerGlyAspSerGlySerProLeuTyr 261
816 TATCATATATGCCCAAAAGCAAAAGTGTTAATTATAGGGGTATTGCAA 865
      ::::|:::
261 rAlaTyrAsnHisPheAsnGlnHisLysTyrValLeuAlaGlyVal... ThrL 277
866 CAGGCAACCCCTATATAGAAAAAGCATGGCTTCCAGCTAGTTCGTAA 915
      |||:::|:::
277 euTyrAsnGlnGlyValAsnGlySerArgAsnTyrPLeuGlnLeuPro 293
916 GATTGGTTCATGATGAAATCTTCTGCTGAGATACCATTCATTCATTTCTA 965
      ::::|:::
294 AsnSerPheIleAsnGlnThrIleLysSerAspSerLysGlnIleValVa 310
966 CGAACCCATCAAAATGGAATACTTTTAAAGACATATATATAGCGC 1015
      ::::|:::
310 LysAsnLysLysGlnAspIleThrTyrSerPheAspLysThrGlnIleT 327
1016 CAGAAAAATGATGCCAAACATAACATATCTTACTTATATATATTA 1065
      ::::|:::
327 hrGlyValIleAsnLeuGlnArgLysAsp..... 336
1066 AAAACAGACACCGTTCAATTGTTAAATGTTCTTATCCGAGACAGAA 1115
336 .....
1116 AGAACCTGTTATCATGCTGCAGGTGGGCTCAACAGTTATGACCCAGAC 1165
      |||
337 .....HistIleIleAsnGlnLysLysAsnAsnAsnPro.... 347
1166 TGAATTAATGAGAAATATTTCTTATTGCAAAAGGAAAGGATGATG 1215
      ::::|:::
348 .AspTyrGlyLysAsnIleValPheSerGlyGlu... GlnGlyValIle 362
1216 ATACTTCACGACATCAACCAAGCGCGCGGCTGTTATTTTGAGGG 1265
      ::::|:::
363 AsnIleLysAsnAspIleAlaGlnGlyPheIleGlnPheAsnAs 379
1266 TAATTTACGGTCTCGCTAAACCAACAGAGGTGCGACAGCGCGCGCG 1315
      ::::|:::
379 nasTyrIleValSerThrGlnAsnAsnSerThrTyrPheGlySerGly 396
1316 TTCATATACGTGATGACGATACCTTACTTGGAAGTAAACGCGCTGGCA 1365
      ::::|:::
396 leIleValAsnAspGlySerGlnValTyrTrpGlyIleAsnGlnValLys 412

```

```

1366 AACGACCGCCTGTCCAAATCGCAAAAGCAACGCTGCTGTTCAAGCAA 1415
      ::::::::::: |||:::
413 AspAspAsnHisLysLysLeuGlySerGlyLysLeuIleIleAsnGlyAs 429
1416 AGGGAAACCAACCAAGCTCGCTGACCGCTGGCGCGGTAAGTATCTTAG 1465
      |||:::
429 nGlyIleAsnThrGlyLeuLysValGlyAspGlyThrValLeuLeuS 446
1466 ATCAGCAGCGCGACGATCAAGCAAAACACCTTTAGTGAATGCGC 1515
      ::::|:::
446 ergIlnHisLysAspAlaGlnGlyArgValGlnAlaPheSerSerValAsn 462
1516 TTGCTGACGGCGAGGGGCGTGCACCTGATAGCCGATATATGATCA 1565
      ::::|:::
463 IleSerSerGlyArgProValIleValLeuSerAspAspLysGlnIleAs 479
1566 CCCCAGACAACTCTATTTCGCTTTCGCGCGGACGCTTGGATTTGAAG 1615
      |||:::
479 nProAspAsnIleSerTyrPglTyrArgGlyGlyThrLeuAspLeuAsnG 496
1616 GGCATTGCTTGTGTTCCACCGCATTCAAATATCCATGAAGGCGGATG 1665
      |||:::
496 LysAsnAsnIleThrPheThrArgIleGlnAlaSerAspTyrGlyAlaIle 512
1666 ATTGTCAACCAATCAACAGACAAAGAA..... 1692
513 IleThrAsnAsnAsnAspLysLysSerIleThrAsnLeuAsnIleSerTh 529
1693 .....
529 rLeuLysAspThrAspIleAsnValSerValSerAspValHisValLeuG 546
1712 GCATTA..... 1719
546 LysGlyLysGlyLysProGlyAspLeuPheTyrAspArgGlyThrGlyArg 562
1720 .....GATATTAC 1727
563 TyrLeuIleLeuLysLysGlnLysTyrSerProPhePheGlnAspValAs 579
1728 TACAC.....GCATTAACAAAC..... 1749
579 nasnThrSerValTyrPglTyrValGlyLysAsnArgAsnGlnAlaIleA 596
1750 .....TTGATAGCAAAAGAAATTCGCTAC 1776
596 lavalValLysSerGlnLysIleLysAsnSerSerLysProTyrLeuPhe 612
1777 AACGTTGGTTGCGGAGAAAGATGCAACAAACAGACGCGCGCTCAA 1826
      ::::|:::
613 HisGly.....GlnLeuAsnGlyAsnMetAs 621
1827 TCGAATTACCAACCGGAAGCGGATCGCTTACTGCTTTCGCGCG 1876
      ::::|:::
621 pIleAsnIleGlnSerGlyMetGlyValAsnThrPheAlaIleAspGlyS 638
1877 GAACAATTTA...AACGCAATATCAGCAACAAACAGGCAAACTGTT 1923
      ::::|:::
638 erValAsnIleProAsnGlnThrLeuSerAsnLysLeuGlySerLeuIle 654
1924 TTCAGCGGACGACGACACCGCGCTACATCATCTTATAGCAAGCGGCTG 1973
      ::::|:::
655 PheGlnGlnHisProValIleHisAla.....GlyValGlyVa 667
1974 GTCAAAATATGAGAGTATCCACAGAGAAATCGTGGGCAACGATT 2023
      ::::|:::
667 lserAlaGlnGln.....AsnAspT 674
2024 GGATGACGCGCATTTAAAGCGGAAACCTTCATATTACAGCGGACAA 2073
      |||
674 rPglThrArgTyrPheThrLeuAsnThrLeuAsnLeuAspLysSerAsp 690
2074 GCGGTGTTCCCGCAATGTTGCCAAAGTGAAGCGCATTTTAAG 2123

```

```
691 pheHisLeuSerArgAsn...AlaValMetThrGlyAsnIleHisAlaThr 706
      :::::::::::::: :::: :::: :::: :::: ::::
2124 CAATCAGCCCAAGCAGATTGTTGGTGTGGACCCGATCAAGCCACACA 2173
      :::::::::::::: :::: :::: :::: ::::
706 rAsnSerSerLysValThrLeuGly.....SerSerSerValT 719
      :::::::::::::: :::: :::: ::::
2174 TCTGTACAGTTGGGACGTGAGGGT.....CTGACAGTTGT 2211
      :::::::::::::: :::: :::: ::::
719 yTValAspArgAsnAspGlyThrGlyGluSerValAsnSerThrSerAla 735
      :::::::::::::: :::: :::: ::::
2212 ACCGAAAAACCATTAACGCGATTAAGTATTGCTTCATTGAGCAAGC 2261
      :::::::::::::: :::: :::: ::::
736 ThrAlaSerAspIleAsnGlu.....LysAspLeuSerLys... 747
      :::::::::::::: :::: :::: ::::
2262 CGACATCAGAGCAATGTACGCTTGGCATGACGCTCATTTAAATCTCA 2311
      :::::::::::::: :::: :::: ::::
748 ...PheThrGlyAspIleThr...AlaAspAsnSerAsnIleIleIleA 762
      :::::::::::::: :::: :::: ::::
2312 CA.....GGACTTGCACACATCAACGCGCAATCTTAGTGCA 2346
      :::::::::::::: :::: :::: ::::
762 snAsnLysPheGluGlyGlyValSerAlaValAsnLysSerAlaIleAsp 778
      :::::::::::::: :::: :::: ::::
2347 GCGGAGACACGCACTATACGGTTAGCGC.....AACGCAACCCAAA 2390
      :::::::::::::: :::: :::: ::::
779 IleHisSerGlnHisAlaValIleAsnArgTrpSerAspIleSerAspAs 795
      :::::::::::::: :::: :::: ::::
2391 CGGCAACCTCAGCCTCGTGGCAATGCCCAACAGACATTTAATCAAGCA 2440
      :::::::::::::: :::: :::: ::::
795 nSerLysLysThrLeuLysLysAsnAlaThrLeuThrValAsnThrGlyL 812
      :::::::::::::: :::: :::: ::::
2441 CATTAAAC...GSCACACATCGGCTTCGACACATGCTTCATTATCTCA 2487
      :::::::::::::: :::: :::: ::::
812 euValAsnLysGlyThrIleGluIleGlyGluAsnAlaGluLeuAsnLeu 828
      :::::::::::::: :::: :::: ::::
2487 ..... 2487
829 GlnGlyTrpProIleAlaAspLysPheIleProSerLeuHisAspLeuG 845
      :::::::::::::: :::: :::: ::::
2488 .....AGCAACAAG 2497
      :::::::::::::: :::: :::: ::::
845 yAsnValLysMetThrAlaSerAsnAlaThrLeuThrAlaGlyAsnTrpA 862
      :::::::::::::: :::: :::: ::::
2498 CCGTACAAACGCGCAGCTGACGCTTTCGACAAACGCTAAGCAACGTA 2547
      :::::::::::::: :::: :::: ::::
862 IamethPheSerGlyGluIleThrAlaAspAspAlaThrAlaValaGlyAl 878
      :::::::::::::: :::: :::: ::::
2548 .....AGCATTCGCGACCTCAACGCGCAAGTGTCTC 2576
      :::::::::::::: :::: :::: ::::
879 AsnLeuGlySerGluThrSerThrLeuSerGluPheAsnProAsnProG 895
      :::::::::::::: :::: :::: ::::
2577 CCTAGCCGATNAGGACGATTCATTTCGAAAAACGCCGTTACCGGAA 2626
      :::::::::::::: :::: :::: ::::
895 uLeuThrAsp...LeuMetPheAspLysTrpAsnThrSerTrpThrGlyL 911
      :::::::::::::: :::: :::: ::::
2627 AAATCAGCGCGGCAAGATAGCATTTACATTAAAGACAGCAGATGG 2676
      :::::::::::::: :::: :::: ::::
911 ystIleSerAlaLeuLysGlyAspAla...SerMetValAsnThrValTrp 926
      :::::::::::::: :::: :::: ::::
2677 ACGTGGCGGTGGGACGGAATTAGGCAATTTAAACCTTGACAAGCGCAC 2726
      :::::::::::::: :::: :::: ::::
927 ArgMetThrGlyAspSerGlyLeuAsnThrLeu..... 937
      :::::::::::::: :::: :::: ::::
2727 CATTAACCTAATTCGCGCTATGACACGATGGGCGGCGCAACCG 2776
      :::::::::::::: :::: :::: ::::
937 ..... 937
2777 GCAATGGGAGATGCGCGCGCGCGCTTCGCGCGCTTCCTATATTC 2826
      :::::::::::::: :::: :::: ::::
938 .....LysThrSerLysSerLeuThrVal 945
      :::::::::::::: :::: :::: ::::
2827 GTTACGCCGCAACTTCGGCAGATCCCGTTTCAACACGCTGACGGTAA 2876
      :::::::::::::: :::: :::: ::::
```


261 rAlaTrpAspAsnIleAspLysLysTrpValLeuAlaGlyValThrLeuH 278
866 CAGGCAACCCCTATATAGAAAACCAATGGCTTCAGTAGTGTAA 915
278 IsAsnYrGlyValLysGlyAlaArgAsnAspTrpLeuIleProHis 294
916 GATTGGTTCTATGATGAATCTTTGCTGAGATACCAATTCATGTA 965
295 AspPheIleSerGlnLysLeuGlnAspAspLeuLysProIleIleValAl 311
966 CGAACCACTCAAAATGGAAATCTTTTAAACACAAATAAATGGCG 1015
311 AserProGlnGluAsnIleLeuArgTrpGluPheAspArgSerArgGlyT 328
1016 CAGGAAAAATCGATGCCAAACATAAAGACTATCTTACCTTATAGATTA 1065
328 hTrGlyThrLeuSerGlnGlyGluLysIlePheSerMet..... 340
1066 AAAACAGAACCGTTCAATGTTTAAATGTTCTTATCCGAGACAGCAAG 1115
340 340
1116 AGACCTGTTTATCATGCTGCAGGCGGTCAACACTTATCCACCAGAC 1165
341ThrGlySerValAsnGly.....AsnA 348
1166 TGAATATGAGAAATATTTCTTATGACAAAGAAAGAGTGAATG 1215
348 IAsnThrGlyAsnAsnLeuValPhe...SerGlyAsnGlnGlyLysIle 363
1216 ATACTTACCAACCAATCAACCAAGCGCGGCGGCTTATTTTTCAGAG 1265
364 GluLeuValSerSerValGlnGlnGlyAlaGlyTyrLeuGlnPheAspLys 380
1266 TAATTTACGGTCGCTCAAAACACAGAAACGTGGCAAGCGCGGCGG 1315
380 sAspTyrThrValLeuThrAsnAsnAsnSerThrTrpGlyAlaGlyI 397
1316 TTCTATTCAGTGTGCGATGACGTTACTGTGAAAGTAAAGCGCGGCA 1365
397 IeIleValGlyAspGlnAlaAsnValLysTrpGlyValAsnGlyIleAla 413
1366 AACGACCGGCTGTCAAAATCGCAAAAGCAGCGCTGTGTTCAAGCAA 1415
414 GlysAspAsnLeuHisLysValGlySerGlyThrLeuThrValAsnGlyI 430
1416 AGGGGAAACCAAGCTCGGTGAGCTGGGCGACGCTAAAGTCATCTAG 1465
430 sGlyGluAsnLysGlyGlyLeuLysValGlyAspGlyValIleValLeuG 447
1466 ATCAGCAGCGGCGACATCAAGCAAAAACAGCTTTAGTGAATCGCG 1515
447 IuGlnGlnProAspAlaAsnGlnLysGlnGlnIlePheSerHisIleAsn 463
1516 TTGGTACGCGGAGGCGAGGTGCAACTGAATGCCGATTAACAGTTCA 1565
464 IleAlaSerGlyArgAlaThrValLysLeuAsnGlyAlaAsnGlnValAs 480
1566 CCCCAGCAAACTATTTGGCTTTCGCGCGGAGCTTTGGATTTGAACG 1615
480 pAlaAspAsnIleSerTrpGlyArgGlyGlyLysLeuAspLeuAsn 497
1616 GGCATTCGCTTGTGTCACCGCATCAAAATACGATGAAGGCGGAGG 1665
497 LysTrpAspPheThrPheSerArgLeuGlnAlaAlaAspTyrGlyAlaGlu 513
1666 ATTGTCAACCAATCAAGCAAGAAAGATCCACCGTTACATTAACAGCAA 1715
514 IleSerAsnAspAsnGlnThrAspLysSerIleValThrLeuSerLeuSe 530
1716 T.....AAGATATTTACTCAACCGCAATTAACACACTTC. 1752
530 rProLeuLysAlaGlnGlnIleAsnValValAlaAsnAsnIleAsnIleW 547

1753GATAGCAAAAAGAAATTCCTCAACAGGCTGTGGC 1791
547 eTrGlyGlyThrGlyLysProGlyAspLeuYrTrpThrPheAspGly 563
1792 GAGAAAGATGCAACCAAAAGAC.....GGCG 1820
564 AsnTrpTyrLeuLeuLysSerAsnArgTyrGlySerAlaLeuPheGlyAl 580
1821 GTCAT.....CTGAATTACCAACCGAGAG 1849
580 aLeuAsnAsnGlnSerGlnTrpGlnArgLeuGlyLysAspLysGlnLysA 597
1850 CG..... 1851
597 IAlleGlyLeuTyrThrGlnMetLysMetGlnIleSerAlaProLeuSer 613
1851 1851
614 TyrIleTyrHisGlyLysIleThrGlyAsnThrSerValGluIleProLys 630
1852GATCGCACTTACTGCTTTCGCGGCAACAAATTTAAACG 1891
630 sLeuAlaGlyAsnAspIleLeuThrLeuAspGlySerValSerIleSerG 647
1892 GCAATATCAGCGCAACAAACGCAAACTGTTTTCACGCGCAGACGACA 1941
647 LysPheMetSerLysGlnAspGlyAlaLeuIlePheGlnGlyHisProVal 663
1942 CCGCAGCGCTCAATCATTTAGAAAGCGGCTGTCAAAAATGCAAGSTAT 1991
664 IleHisAlaGlyGlnThrValSerAlaSerGlnSer..... 675
1992 CCCACAAAGAAATTCGTGTGGACACGATGGATGATCGACCGACATTTA 2041
676AspTrpGluAsnArgLubPheS 683
2042 AAGCGAAACCTTCATATTCAGGCGGACAAAGCGGTGTTCCCGCAAT 2091
683 eTrLeuAsnAsnLeuAsnLeuAsnAsnAlaAspPheSerLeuSerArgAsn 699
2092 GTTGCCAAAGTGGAGGCGATTTGCAATTAAGCAATCAACGCCCAAGCA 2141
700 ...AlaPheMetAsnGlyAsnIleArgAlaValAsnGlnSerThrValI 715
2142 TTTCGGTGTGCGACCGCATCAAGCGCANACATCTGTACA...CGTTCG 2188
715 eIleGly.....GlyAspThrValPheThrAspLysAsnA 727
2189 ACTGACGCGGCTG..... 2202
727 sPcGlyThrGlyAsnAspValIleSerValGlnGlyLysSerAlaAlaAla 743
2203 ...ACAAGTTGACCGAAAAAACCATTAACGACGATTAAGATTTGCTTC 2249
744 GlyThrSerSerTyrThrGlyHisIleThrLeuGlnGlnLysSerAlaLe 760
2250 ATTGACGCAAGACCGCATCAGAGGCAATGTCACCTTCGATCAGCGTC 2299
760 uAspIleArgAspAsnPheArgGlyGlyValThrSerGluAsp...SerH 776
2300 ATTTAAATCTCAGAGACTTCCACACTCAACGCGCAATCTTACGACGCG 2349
776 IsIleAsnValSerSerSerSerValLeuPheSerAspAlaSerSerPhe 792
2350 GGAGACGCGACTATACGTTACGCGCAAGCC.....ACCCAAACGG 2393
793 IleAsnSerSerLeuAsnIleHisLysGlyAlaLeuThrAlaGlnGlnI 809
2394 CAACCTCAGCGCTGTTGGCAATGCCCAAGCAACATTAATCAAGGCACAT 2443
809 yGlyLeuPheThrSerGlySerIleAspIleGlyAspAlaSerLeuLeuL 826

```

2444 TAAAGCAACATCGCTTCGGACATGCTTCAATTAATCAAC 2493
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
826 eutrhgltlhrProValAsnSerAspAlaIleLeuProthrlle 842
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2494 AACGGCTCAAAAGGAGGTGAGCGCTTCCGACACGCT....AA 2537
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
843 AsnMetAlaSpIeYlPheYlSLeuMetSerAspSerValLeuY 859
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2538 GGCACAGTAAGCCATTCCGACACTCAACGCAATGCTCCCTAGCCGATA 2587
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
859 salaary...AspGlnAlaSerValValGlySplIleIleSerAspYsg 875
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2588 AGGCAGTATTCATTTTGAACAGCGCGCTTACCGGAAAAATCAAGCGCG 2637
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
875 lnaIatlrIleSerPheGluThr.....GluSer 884
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2638 GGCACAGTACGGCATTTACACTTAATAAGACAGGAATGACGCTCCGCG 2687
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
885 GlyYsgIleGlyMetLeuSerGluYlSAlaSerArg..... 896
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2688 GGGCAGCGAATTAGGCAATTTAAACCTTGACAAGCCACCATTTACATCA 2737
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
897 .GlyLeuAlaValAlGlyLeuLeuSer.....GlyPheA 907
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2738 ATTCGCCCTATGACACAGATCGCGCAGCGCGCAACCGGACGTGGCGCA 2787
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
907 smthralatYarg.....GlyAlaIleHisAlaProSerAla 919
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2788 GATGCCCGCGCGCGCGCT.....TCGCG 2810
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
920 SerAlaIleMetAsnAsnThrTrpTrpGlnLeuThrGlyAspSerSerle 936
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2811 CGGTTCCTATTATCGGTATACCGCGCACT.....TCG 2845
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
936 uArgSerLeuYlSAsnThrGlySerMetThrYrPheThrGlySerAla 953
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2846 CAGAAATCCGCTTCAACAGCTGAGGCAATTAAGGCAATGACGCTAG 2895
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
953 lAsnYlSAlaPheHisThrLeuThrValAsp...GluLeuThrThrSrn 968
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2896 GGAACATTCGCGCTTATGTCGGAACCTTCGCTACCGCAGCGGCAAAAT 2945
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
969 GlyThr.....AlaYrAlaMetAlaYthr..AspLe 978
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2946 GAACTGGCGGAAAGTCCGAAAGCACTTACCTTGCTGTCACAATA 2995
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
978 uYlSAsnAlaAspYlS.....LeuValValAsnGlnL 989
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
2996 CCGGCAAGCAACCCGTAAGTCTCGAGCAATTGACGTAAGTGAAGAGAAA 3045
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
989 Ys.....LeuSerGlyYlS 993
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3046 GACAACACACGCTGTCGAAATCTTAATTACCTGCAAAAGCA 3095
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
994 AspAsnIleLeuLeuValAspPheLeuSrn..... 1003
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3096 CGTCAATGCGCGCGCATGCGCTTATCACTTATCCGAAAGACGCGAGT 3145
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1003 ..... 1003
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3146 TCCTCGCTCATTAATCCGCTCAAAAGACAAGCTTTCGACAACTCGGC 3195
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1003 ..... 1003
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3196 AAGGCGGAGAAACAGAGCGCGCTTGACGCAAAACAGGCAACATTGC 3245
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1003 ..... 1003
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3246 CGCCAAACACAGGCGGAAAAAGCAACGCGCAAGCGCTTGAC...GCGC 3292
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1004 .....LysProThrGlyGlyLysLeuSrnPlleGlnL 1014
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3293 TGATTGGCGCGCGCGCATGCGCACCGAAAGCAAGAAAGTGTTCGCGAA 3342
      |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1014 eutValSerAlaProGlyLysnSerSerYlSAspValIlePheYlSgYserGlu 1030
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3343 CCGGCGCGCGCAGGACGCGGGGAAAAAT...GCGCGCATTTATGACGCGCA 3389
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1031 GlnAla.....IleGlyPheSerAsnValThrProValIleThrAlaY 1045
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3390 GGAAGAGAAAAACGGGTGCGAGCGGATTAAGACACC...GCTTGCGCA 3436
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1045 sAsp.....AlaGlyAspYlSThrThrTrpAsnLeuThrG 1057
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3437 AACAGCGGAAACGGAACCGCGCGCTTACACCGCGCTTCCCGCGCC 3486
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1057 LyTyArgMetAlaGluAsnProAlaIleThrGlnSer..... 1069
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3487 CGCGCGCGCGCGCGGATTTGCCCAACCGACCCCAACCGCAACCCCA 3536
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1069 ..... 1069
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3537 ACCGACGCGGACCTGATCAGCCGTTATGGCAATAGCGTTTGAGTGAAT 3586
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1070 .....AlaSerGlyLeuAlaSerValGlyTyIlylSerP 1081
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3587 TTTCGCCACGCTCACAGCGGTTTTCGCGGTACAGACGAATTGACCGC 3636
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1081 heLeuSerGluValAsnAsnLeu..... 1088
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3637 GTGTTGCCAGAACGCCCGCCACGCGGTTTGACAAAGCGGCATCCGGCA 3686
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1088 ..... 1088
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3687 CACCAACACTACCGTTCGCAAGATTTCCGCGCTTACCGCAACAAACG 3736
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1089 .....AsnLyArgMetGlyA 1094
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3737 ACTGCGCCAAATCGTATGACGAAAAACCTC..... 3768
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1094 splEuAdgAspIleAsnGlyGluAlaGlyAlaThrAlaArgIleMetSer 1110
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3769 GGCAGCGGCGCGCGCGCATCCTGTTTTCGCAACACGCGGCAACAC 3818
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1111 GlyThrGlySerAlaGlyGlyGlyPheSerAspAsnHisThr.....Hi 1125
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3819 CTTCGACGAGCGGATCGGCACTCGGACGCGCTTGCCACGCGTCCGTT 3868
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1125 sValGlnValGlyValAspYlSAsnGlyLeuAspArgGlyLeuAspLeuP 1142
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3869 TCGGGCAATACGCGATCGGCACTCGGATCGGATCGACGCG..... 3912
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1142 heThrGlyPheThrValThrHisThrAspSerSerAlaSerAlaAspAla 1158
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3913 .....GGCGCGGCT..... 3921
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1159 PheYlSgYlSThrYlSserValGlyAlaGlyLeuThrYlSAlaSerAlaMe 1175
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3922 .TTTAGTAGCGGCGCGCTTTCAGACGCGCATCAAGGCAAAATCCGCGCC 3970
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1175 tPheAspSerGlyAlaTyIleAspLeuIle...GlyYlS..... 1187
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
3971 GCGTGCATTCAGCGCATTCAGGCAAGATTCGCGCGAGTTTCGCGGGA 4020
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1188 ..TyIValHisHis.....AspAsnGluTyIThrAlaThrPheAlaGly 1201
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
4021 TTCGCGATCGAAGCGGACATCGGCGGCAAGCGGTATTGTGTCCAAAAC 4070
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1202 LeuGlyYlLeuArg..... 1205
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
4071 GGAATACCGATACGAAACGTCATATCGCACACCGCGGCTTGCAATCA 4120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1206 .AspTyI.....SerThrH 1210
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
4121 ACCGTAACGCGCGCGCATTAAGGCAATTAATTCATTAACCGCGCGCA 4170
      |||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

298 ...ileGlySerThrAlaValArgLeuAlaAsnGluArgAspAlaAs 313
1170 TATGAGAGAAATATTTCTTATTGACAAAGCAAGCAATGATGATAC 1219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 nasnglyGlnAsnValThrPheGluAsp.....AsnGlyThrLeuVal 328
1220 TTACCAACACATCAACCAAGCGCGGGGCTTTGATTTGAGGTAAT 1269
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 euAspGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheGlyGlyAsp 344
1270 TTTAGCGTGCCTAAACAAACGAA...ACGTGGCAAGCGCGGGCT 1316
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 TTTThrValGlyAlaAsnAsnSpiriThrPleuGlyAlaGlyI 361
1317 TCATATCAGTATGAGCAGTACCTGACTGTGAAGTAAACGCGCGGCA 1366
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 euAspValAlaAspGlyLysLysValValTropInValLysAsnProAsnG 378
1367 ACGACCGCTGTCCAAATCGCAAGGACGCTGCTGTTCACGCAAA 1416
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 LysAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThr 394
1417 GGGGAAACCAAGCGCTCGAGCGTGGCGGCGGTAAGTCATCTTGA 1466
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 GlyValAsnGlnGlyLeuLysValGlyAspGlyThrValIleLeuAs 411
1467 TCAGCAGCGGACGATCAAGCAAGCAAGCCTTAGTGAATCGGCT 1516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 nGlnGlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyI 428
1517 TGGTACGCGGCGGAGCGGTGCACTGATCCGATATCAGTTCAAC 1566
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
428 LeuAlaSerGlyArgGlyThrLeuValIleuAsnSerSerAsnGlnIleAsn 444
1567 CCGGCAACCTATTTGCGGCTTGGGAGCGTGGATTTGAACGG 1616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
445 ProAspAsnLeuThrPheGlyPheArgGlyGlyArgLeuAspAlaAsnG 461
1617 GCATTGCGTTGCTCCACCGCATCAAAATACGATGAAGCGGATGA 1666
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
461 YAsnAspLeuThrPheGlnHisIleArgAsnValAspGlnGlyAlaArgI 478
1667 TTGTACACCAATCAAGCAAGCAAGATCCCGTTACATTACAGGCAAT 1716
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
478 LeuAlaAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGlyLys 494
1717 AAAGATTACTACACCGGCAAT 1740
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 SerLeuIleThrAlaProGlnAsn 502

seq_name: sp_bacteria:051165

seq_documentation block:
ID 051165 PRELIMINARY; PRT; 496 AA.
AC 051165;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteri: Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG44/76;
RX MEDLINE=95302961; PubMed=7783620;
RA Lombolt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding Iga1 protease
in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
influenzae";
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82481; CAA57864.1; -.

```

```

DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASEPRPASE.
KW Protease.
FT NON_TER 1
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 496 AA; 54645 MW; AED2478E9F3CB84 CRC64;

alignment_scores:
  Quality: 875.50 Length: 551
  Ratio: 2.459 Gaps: 20
Percent Similarity: 64.610 Percent Identity: 38.294

alignment_block:
US-09-303-518D-653 x 051165 ..

Align seg 1/1 to: 051165 from: 1 to: 496

202 CGGAAAGATATTTGAGTTTACAAACAAAAGGAGCTGTGCGCAATC 251
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 GATGACGCAAGCCCGATGATGATTTCTGTGTATCGCTAAC...G 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 AlaSerAsnValAlaProMetIleAspPheSerValAlaAspValAsnArg 34
299 GCGTGGCGGATTTGGCGGCGGATCAATATTTGTAGCGTCACATAC 348
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 rGlyThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50
349 GCGGGC.....TATACAAATGTTGATTTGGCGGAGGG 383
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 LysGlyAspSerLulleSerTyrGlyGlnHisAsnGlnHisLeuAsp 67
384 AAGCAATCCGATCAGACCGCTTTCTTACCAAAATTTGMAAAGAAATA 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 LserAsnAspGluAsn.....GluTyrArgSerValAlaGlnAsn 81
434 ATTATTAAGCAGGAGACTAACGCCATCCTTATGCGCGC..... 471
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 sPyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeu 96
472 ..GATTATCATATGCGCGTTTGCACAAATTTGCACAGATGCCAAGC 518
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 GluAspTyrAsnMetAlaArgLeuAsnLysSpreValThrGluValAlaPr 113
519 TGTGAGATGACCGATTATATGATGATGGTGAA...TACGCTGATTGA 565
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 oIleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLys 130
566 ATTAATACCTGATCGTTCGATCGAGACGACGACATATTTGGCGG 615
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
130 snArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe..... 144
616 TCGATGAAGAGACCAACCAATACCGCGAAAGTTCATATCATATTCAG 665
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145 .....GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157
666 CGCATATTCCTTGGCTGCTGCGTGCATATCTTT.....G 700
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 gAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174
701 CACAAATGATGAGTGTGTCGACAGTCACTTACGATGACGCAAAAT 750
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 hSerAsnLeuAsnGlnGlnGlyLeuIleGlyPheGlyLysPAsnSerLys 190
751 AAACATATGCGCA.....TANGTTTTTAC 776
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 HisHisSerProGluLysLeuLysGluValIleuSerGlnAsnAlaLeuTh 207
777 AACAGAGGCTCATTTGGCAGACAGTGGCTACCAATGTTATCTATGATG 826

```


67 1serfasmnspgluasn.....glutyrargsetvalalaglinsn 81

434 attataaagcaggacntracgccatcttatgccc..... 471

81 sptrygluiproasnlysntrphr...hisglyasnngluyatrgleu 96

472 ...gattatcatatgcccgggttgcacaaattgtctcacatgacagacc 518

97 gluaspryasnmetlalaigleasnlylphelvalthrcgluvalalapr 113

519 tcttgacatgaccgcttatatgctggcgaaa...tacgctgatttaa 565

113 oilealaprthrserlaiglygllyvalgluttryltylaspyaspya 130

566 ataataaccggatcgatgggtaacggacagacagacaaatrtggcgg 615

130 snatrgpasesgluphelaatrgvalajalaglylthrglnphe..... 144

616 tctgatgaagacgaacccaataaccgggaagtctcatatcatatgacag 665

145glutyranserargtyasnmethcgluensear 157

666 gcgatattcttgctgcgtgcgaatraccttt.....g 700

157 galatryargtyralallelaaglythrprotyrglnaspyalasnvalt 174

701 cacaaaatgcatgagctggtgcacagctcaactyagtagcggaaaaatt 750

174 hrserasnleasnngluclylguileuilelylphedylaspsnseryls 190

751 aaacatgacca.....tatgtyttttacc 776

191 hishiserprrogluylsleuylsgluvalleuSERglinsalaleuth 207

777 AACGAGGCGCTCATTTGGCACAGTGCATCCAAATGTTATCTATGANG 826

207 rasnttyalavalencllyaspereltyserproleuphelaatryaspl 224

827 cccaaaagcaaaagctgatttaattgaggatgttgcamaacagcacc 876

224 ysglngluysatrgtrpvalphelenglyalatyraspytrpralagly 240

877 tatatagaaaaagcattggcttccacagctagttcgtaaaagtgg...tt 923

241 tyr.....gluysasnsertrpgln.....glutyrasnll 251

924 ctatgatgaatcttgcctgagatgaccattcagatattctacgaaccac 973

251 etryllyslsgluaphelaasplguilelysgln..... 262

974 atcaaaatgggaatnactttttaacgacaatratatggccagagaaa 1022

263atgaspasn.....alaglythr 268

1024 atcgatgccaaacatatacactattcttaccctttagattaaaaacg 1072

269 llelysglylrglygluhishs.....trplythtrth 280

1074 AACCGTTCAATGTTAAAGTTCCTTATCCGACGACGACGAGAACCCTG 1122

280 rgllythr.....asnserhisllelyserthrhalavalalaglea 294

1124 tttatcatgctcgagctggggctcaacagtattcgaaccgactgaatat 1173

294 laelylsnngluatrglyala.....asnsn 302

1174 GCGAATAATATTTCTTATATGACAAAGAAAGGTGAATTCATCTTAC 1223

303 glyglinsnvalthrphe.....gluasnasnnglythrleuvalleuas 317

1224 CACCAATCAACCAAGCGCGGGGTTGATTTATTTAGGGTAAATTTA 1273

317 pglinsnllleasnnglueylalaglylguileupheheylslyaspyrrt 334

alignment_scores:

Quality: 864.50 Length: 536
 Ratio: 2.449 Gaps: 20
 Percent Similarity: 65.858 Percent Identity: 38.993

alignment_block:

US-09-303-518D-653 x QS1163

Align seg 1/1 to: QS1163 from: 1 to: 496

```

202 GCGAAGATATTTAGGTTTACACAAAAAGGAGTTGTCGGCAATC 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 AlatrhaepleuserVallysasnlysglnlglnasnlllelglyAsnAl 17
252 GATGACGAAGCCCGCATGATTGATTTTCTGTGTGATCCGCTAAC...G 298
    : : : : : : : : : : : : : : : : : : : : : : : :
17 aleuserasnValPrometileaspheserValAlaAspValAsnArgA 34
299 GCGTGGCGCATTTGGCGGCGCATATATTTGAGCGGCACATTAAC 348
    : : : : : : : : : : : : : : : : : : : : : : : :
34 rghThleuthValIleasproglntyrAlaValSerVallyshHisVal 50
349 GCGGCGC.....TATAACATGTGATTTTGGTCGCGAGG 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 lysglyAspGluIleSerTyrTgIyHshIsasnlglyHlsleuAspVa 67
384 AAGCAATCCCGATCAGACACCGCTTTTCTTACCAATTTGTGAAGAATA 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 lserasnspGluasn.....GluTyrArgSerValAlaGlnAsnA 81
434 ATTATAAAGCAGGACTTAACGGCATCTTATGCGGCGC..... 471
    : : : : : : : : : : : : : : : : : : : : : : : :
81 spTyrGluProasnlysasntRphIs...HisGlyAsnlglyArgleu 96
472 ...GATTATCATATCCCGGTTTGCACAATTTGTACAGATGACGAAC 518
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
97 GluAspTyrAsnMetAlaArgleuasnlyspheValThrGluValAlaPr 113
519 TGTGATGATGACCACTTATATGATGGTGGGAAA...TACGCTGATTAA 565
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 otleAlaProThrSerAlaGlyGlyValGluThrTyrIlyAspLysA 130
566 ATAAATACCCTGATCGTGTTCGATCGAGCAGCAGACAATATATGGCGG 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 snArgpheserGluPheValArgValGlyAlaGlyThrGlnPhe..... 144
616 TCTGATGAAGACGAACCCATTAACCGCGAAGTTCATATCATATTTGCAAG 665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 .....GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157
666 CCGATATTTTGGCTCGTCGCGCATTAACCTTT.....G 700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 galatyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174
701 GCAAAATGATCAGGTGGTGGCAGACAGTCACTAGTAGCGAAAAAAT 750
    : : : : : : : : : : : : : : : : : : : : : : : :
174 hrSerasnleuasnlglnlglyleuIleGlyPheGlyAspAsnSerIys 190
751 AACAATAGCCCA.....TATGCTTTTATAC 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 HshIsSerProGluIlyleuIySgluValLeuSerGlnAsnAlaLeuTh 207
777 AACAGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCATATGATG 826
    : : : : : : : : : : : : : : : : : : : : : : : :
207 rAsnTyrAlaValleuGlyAspSerGlySerProleuPheAlaTyrAspL 224
827 CCGAAGAAGCAAAAGTGTTAATTAAGGGTATTCGAACAGACGAACCC 876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 ysglnGlyIlyAsnTyrValPheleuGlyAlaIlyAspTyrTrpAlaGly 240
877 TATATGAGAAAAAGCAATGCTTCACAGTACTGTAAGATTCG...TT 923
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

241 Tyr.....GlnIySasnSerTrpGln.....GluTrpAsnIl 251
924 CTATGATGAATATCTTCTGAGATACCATTCAGTATCTACAGAACCC 973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 eTyrIlySlysglnPheAlaAspGluIleIySgln..... 262
974 ATCAAAATGGGAATACTTTTAAAGCACAATTAATGAGCGCAGCAAAA 1023
    : : : : : : : : : : : : : : : : : : : : : : : :
263 .....ArgAspAsn.....AlaGlyThr 268
1024 ATCGATGCCAACAATAAACACTATTTCTTACCTTATAGATTAAACACG 1073
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 IleSlysglyasnlglnIyHshIs.....TrpIySThr 280
1074 AACCGTTCATTTGTTATGTCTTCTTATCCGACAGACAGCAAGAACCG 1123
    : : : : : : : : : : : : : : : : : : : : : : : :
280 rGlyThr.....AsnSerHisIleGlySerThr..... 289
1124 TTTATCATGCTGACAGGTGGGTCACAGTATTCGACCCACACTGAATAT 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
290 .....AlaValArgleuAlaAsnAsnlglnArgAsp...AlaAsnAsn 302
1174 GGAGAAATATTTCTTATTTGACAAAGCAAAAGGTGAATGTACTTAC 1223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
303 GlyGlnAsnValThrPhe.....GluAsnAsnGlyThrIleuValIleuAs 317
1224 CAGCAACATCAACCAAGCGGCGGCTTGTATTTTGAAGGTAAATTTTA 1273
    : : : : : : : : : : : : : : : : : : : : : : : :
317 pGlnAsnIleasnlglnIyAlaGlyGlyLeuPhePheIySglYAspTyr 334
1274 CGGTCTCGCTTAACAAACAGCA...ACGTGGCAAGCGGCGGCTTCAT 1320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
334 hrValIySgIyIleasnAsnAspIleThrTrpIleuGlyAlaGlyIleAsp 350
1321 ATCAGTATGCGACGTACCGTACTTGTGAAGTAACGCGCTGGCAACGA 1370
    : : : : : : : : : : : : : : : : : : : : : : : :
351 ValAlaAspGlyIySlySValValThrGlnValIySasnProasnGlyAs 367
1371 CCGCTGTCGCAAAATGCGCAAGGACGACGCTGCTGCTCAAGCCAAAGGG 1420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 pArgleuAlaIySIlleGlyIySgIyThrIleuGluIleasnGlyThrGly 384
1421 AAAACAGAGCTCGCTCAGCGTGGCGACGCTAAAGTCATCTTAGATCAG 1470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
384 alaAsnlglnGlyIleuIySValGlyAspGlyThrValIleleuAsnGln 400
1471 CAGCGGACGATCAGCGCAAAACAAAGCCTTTAAGTAAATCGGCTTGT 1520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 GlnAlaAspAlaAspIySlySValGlnAlaPheSerGlnValGlyIleVa 417
1521 CAGCGGACAGGCGGAGCGTGCACACTGAATGGCGAATCATGTAACCCCG 1570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 lSerGlyArgGlyThrIleuValIleuAsnSerSerAsnGlnIleasnTroA 434
1571 ACAAACCTATATTTGCGCTTCCGCGGCGAGTGTGATTTGAACGCGCAT 1620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
434 spAsnIleTyrPheGlyPheArgGlyGlyArgleuAspAlaAsnGlyAsn 450
1621 TCGCTTTCGTTCCACCGCATTCAAATACCGATGAGGCGCATGATGT 1670
    : : : : : : : : : : : : : : : : : : : : : : : :
451 AspIleuThrPheGlnHisIleArgAsnValAspIleuGlyAlaArgIleVa 467
1671 CAACCAATGAAGCAAAAGATCCACCGTATACCATTAAGGCAATTAAG 1720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
467 lAsnHisasnThrGlyHisAlaSerThrIleThrIleuThrGlyIySerL 484
1721 ATATTACT 1728
484 euIleThr 486

```

seq_name: sp_bacteria:QS1166

seq_documentation_block:


```

384 alaSnGInGlyInLeuLysValGlyAspGlyThrValIleLeuAsnGln 400
1471 CAGGGGAGAGCTACAGCAAAAGAGCGCTTAGCAAAATCGCGTGGT 1520
401 LysAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyLeuA 417
1521 CAGCGGAGGGGAGCGTGCACATGAAATCCGATATACAGTCAACCCCG 1570
417 lSerGlyArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProA 434
1571 ACAACTGATTTTCGCTTCGCGGCGGACGTTGGATTGACGGGCAAT 1620
434 spAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 450
1621 TCGCTTCGTTCCACCGCATTCGAATACCGATGAAGGGCGATGATTGT 1670
451 AspLeuThrPheGlnHisIleArgAsnValAspGlnGlyAlaArgIleVa 467
1671 CAACCAACAATCAAGACAAGAAATCCACGTTACATTAACAGCAATAAAG 1720
467 lAsnHisAsnThrGlyHisThrSerThrIleThrLeuThrGlyLysSerL 484
1721 ATATTACTACACCGGCAATACACACTTG 1752
484 euIleThr.....AsnProAsnSerLeu 491

```

seq_name: sp_bacteria:Q51167

seq_documentation_block:

```

ID Q51167 PRELIMINARY; PRF; 508 AA.
AC Q51167;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IGAL PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NGC16;
RX MEDLINE=95302961; PubMed=7783620;
RA Lombolt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding IGAL protease
in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
influenzae.";
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82485; CAA57868.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGAL; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 508
SQ SEQUENCE 508 AA; 55879 MW; 9A8BD76477BA25C0 CRC64;

```

alignment_scores:

```

Quality: 859.50 Length: 547
Ratio: 2.463 Gaps: 18
Percent Similarity: 63.803 Percent Identity: 38.391

```

alignment_block:

US-09-303-518d-653 x Q51167

Align seg 1/1 to: Q51167 from: 1 to: 508

```

202 GCGAAAGATTTGAGCTTACACAAAGAGGAGTGTGCGCAATC 251
||| |||.....||| |||.....|||.....: |||.....:
1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlnAsnAl 17
252 GATGACGAAAGCCCGATGATGATTTTCTGTGATGCCGTAAAC...G 298

```

```

.....: |||.....|||.....|||.....: |||
17 aleuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34
299 GCGTGGCGGCAATGCGGGCGATCAATATATTGTCAGCGGCACAT... 345
.....: |||.....|||.....|||.....: |||
34 rGleAlaThrValValAspProGlnThrAlaValSerValLysHisAla 50
346 .....: |||.....|||.....|||.....: |||
51 LysAlaGluValHisThrPheTyrGlyGlnThrLysAsnGlyHisAsnAs 67
372 TGGTGGGAGGAGGAACAATCCCGATCAGACCGCTTTCTTACCAATTG 421
|||.....: |||.....|||.....: |||
67 pValAlaAspLysGlnAsn.....GlnTyrArgValY 78
422 TGAAGAAGAAATATATTAATTAACGAGGACTAACGGCCATCCCTATGCGG 471
|||.....: |||.....|||.....: |||
78 aIGlnGlnAsnAsnTyrGlnPro.....HisLysAlaTyrGly 90
472 .....: |||.....|||.....|||.....: |||
91 AlaSerAsnLeuGlyArgLeuGlnAspTyrLysMetAlaArgPheAsnLy 107
498 ATTGTGCACAGATGCGAGACCTGTTGAGATGACCATTAATGATGGCT 547
|||.....: |||.....|||.....: |||
107 spHeValThrGlnValAlaProIleAlaProThrAspAlaGlyLysL 124
548 GG...AATACGCTGATTAATTAATACCTGATCCGTGTCGAATGGA 594
|||.....: |||.....|||.....: |||
124 euAspPThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140
595 GCAGGACAGACAA.....TATTGGCGCTGATGAAGACGAACCAATAA 638
|||.....: |||.....|||.....: |||
141 AlaGlyArgGlnLeuValTyrGlnLysGlyAlaTyrHisGlnGlnLys 157
639 CCGGCAAGTTCATAT.....CATATTGCAAGCGCATATTCTTGGC 679
|||.....: |||.....|||.....: |||
157 n...GlnLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrA 173
680 TCGTGGGTGGC.....: |||.....|||.....: |||
173 lAlaIleAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsn 189
694 ACCTTGCACAAATGATGATCGAGTGTGGCACAGTCACTAGTAGGCA 743
|||.....: |||.....|||.....: |||
190 ThrGlnGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaG 206
744 AAAAAATTAACATAGC...CCATATGTTTTCACCAACAGAGGCTCAT 790
|||.....: |||.....|||.....: |||
206 uGlnLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValL 223
791 TTGGGAGACAGTGGCTCACCAGTGTATCTATGATGCCCAAGCAAG 840
.....: |||.....|||.....: |||
223 euGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 239
841 TCGTTAATTAATGAGGTATTCGAACAGGCAACCCCTATATAGCAAAAG 890
|||.....: |||.....|||.....: |||
240 TrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLysLysE 256
891 CAATGCTTCACGATAGTTCGTAAGATTTGCTTATGATGAATCTTTG 940
|||.....: |||.....|||.....: |||
256 TrpGlnGlnTyrAsnIleTyrLysLysGlnPheAlaAspLysIleLysG 273
941 CTGGAGATACCATTCAGATATCTCGAACCACATCAAAATGGGAATAC 990
|||.....: |||.....|||.....: |||
273 lArgAsp.....: |||.....|||.....: |||
991 TTTTAAAGACATATATATGCGCGAGAAATATGATGCCAATATA 1040
|||.....: |||.....|||.....: |||
276 .....: |||.....|||.....: |||
1041 ACAGTATTCTACCTTATAGATTAAAAACAGAACCGTCAATTTGTTTA 1090
|||.....: |||.....|||.....: |||

```



```

191 HSHISerProGluLysLeuYsgLValLeuSerGlnAsnAlaLeuH 207
777 AACAGAGGTCATTTGGGACAGTGGCCACCAATGTTATCTTATGANG 826
207 rAsnTYrAlaValLeuGlyAspSerGlySerProLeuPheAlaTYrAspL 224
827 CCCAAAGCAAAAGTGTAAATTAATGAGGTATTCGAAACAGCAACCCC 876
224 ysgLInLysArgTrpValPheLeuGlyAlaTYrAspTYrTrpAlaGly 240
877 TATATAGGAAAAAGCATGGCTTCACGCTAGTTCGTAAGATTGG...TT 923
241 TYr.....GlnYsAsnSerTrpGln.....GluTrpSniL 251
924 CTATGATGAATCTTGTGCGAGATACCATTCAGTATCTTCAGAACCC 973
251 eTYrLysLysGlnPheAlaAspGlnLLeuYsgIn..... 262
974 ATCAAAATGGGAATACTTTTAAACGACAAATATGAGCGCAGGAAAA 1023
263 .....ArgAspAsn.....AlaGlyTYr 268
1024 ATGATGCCAAACATTAACATATCTGTAACCTTATAGATTAAACACG 1073
269 lLeLysGlyTYrGlyGlnHSHs.....TrpLysThr 280
1074 AACCGTTCAATGTTAATGTTCTTCTTATCCGAGACAGCAAGAACCG 1123
280 rGlyTYr.....AsnSerHislLeGlySerTrpAlaValArgLeuA 294
1124 TTATATCATGCTGCAGGTGGGTCAACAGTATATGACCCAGACGATTAAT 1173
294 lAeLysAsnGlnArgGlyAla.....AsnAsn 302
1174 GGAAGAAATATTTCTTATTTGACAAAGAAAGTGAAATGATCTATAC 1223
303 GLyGlnAsnValThrPhe.....GluAsnAsnGlyThrLeuValLeuAs 317
1224 CAGCAATCAACCAAGCGCGCGGCTTGTATTTGAGGTAATTTTA 1273
317 pGlnAsnAlLeAsnGlnGlyAlaGlyLeuPhePheLysGlyAspTYr 334
1274 CGGCTCGCCTTAAACAAACGAA...ACGTGCAAGCGCGGCGCTCAT 1320
334 hrValLysGlyAlaAsnAsnGlylLehrTrpLeuGlylAeGlylLeAsp 350
1321 ATCAGTATGAGCAGTACCTTACTTGGAAAGTAACGGCGTGGCAACGA 1370
351 ValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 367
1371 CCGGCTGCCAAATCGCAAGGCAAGCAGCGCTGTTCAACCAAGGGG 1420
367 pArgLeuAlaLyslLeGlyLysGlyThrLeuGlnlLeAsnGlyThGlyL 384
1421 AAAACCAAGGCTCGTACGCGTGGCGACGTAAGTCAATCTAGATCAG 1470
384 alAsnGlnGlyGlnLeuLysValGlyAspGlyThrVallLeLeuAsnGln 400
1471 CAGCGGACATCAAGGCAAAACAAAGCTTATAGTGAATCGGCTTGGT 1520
401 GlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyLeVa 417
1521 CAGCGGAGGCGGACGTCGACGATGATGCGGATATCACTTCACCCCG 1570
417 lSerGlyArgTrpThrLeuValLeuAsnSerProAspGlnlLeAsnProA 434
1571 ACAACTCTATTTGCGCTTCGCGCGAGCGATTTGATTGAACGGGCAT 1620
434 snAsnLeuTYrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 450
1621 TCGCTTTCGTTCCACCGCATTCAAATACCGATGAAGGGCGCATGTTGT 1670
:::|||||:::

```

```

451 AspleuThrPheGlnHSHsLLeArgAsnValAspGlnGlyAlaArgLeVa 467
1671 CAACCACATCAAGCAAAAGAAATCCACCGTTCATTCATTCAGCAATAAG 1720
467 lAsnHSHsThrAspArgAlaSerThrLeuThrLeuGlyLysSerL 484
1721 ATATTACTACCAACCGGCAAT 1740
484 euLeThrAlaProGlnAsn 490

```

seq_name: sp_bacteria:Q51172

seq_documentation_block:
ID_Q51172 PRELIMINARY; PRT; 496 AA.

```

AC_Q51172;
DT_01-NOV-1996 (TrEMBLrel. 01, Created)
DT_01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT_01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE_IGAI PROTEASE (FRAGMENT).
GN_IGAI.
OS_Neisseria meningitidis.
OC_Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX_NCBI_TaxID:487;
RN_[1]
RP_SEQUENCE FROM N.A.
RC_STRAIN=NGC80;
RX_MEDLINE=95302961; Pubmed=7783620;
RA_Lomholt H., Poulsen K., Mogens K.;
RT_Comparative characterization of the iga gene encoding Iga1 protease
RT_In Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT_Influenzae.
RL_Mol. Microbiol. 15,495-506(1995).
DR_EMBL: X82479; CAA57862.1; -.
DR_MEROPS: S06.001; -.
DR_InterPro: IPR000710; IGA_S6.
DR_Pfam: PF02395; IGA1; 1.
DR_PRINTS: PR00921; IGASERPTASE.
KW_Protease.
FT_NON_TER 1 1
FT_NON_TER 496 496
SO_SEQUENCE 496 AA; 54865 MW; BEFC7A6003FE5D88 CRC64;

```

alignment_scores:
Quality: 856.50 Length: 536
Ratio: 2.433 Gaps: 20
Percent Similarity: 65.672 Percent Identity: 38.806

alignment_block:
US-09-303-518D-653 x Q51172 ..

Align seg 1/1 to: Q51172 from: 1 to: 496

```

202 GCGAAGATATTGAGCTTTACACAAAAAGGAGAGTGGTCGCAATC 251
||| |||:::||||| ||| |||:::|||||:::
1 AlatrhrAspleuSerValLysAsnGlyGlnGlyGlnAlaLysVal 17
252 GATGACGAAGACCCGATGATGATTTCTGTGTATCGGTAC...G 298
:::|||||:::||||| ||| |||:::|||||:::
17 AlAuSerAsnValPrometlLeAspPheSerValAlaAspValAsnAlaG 34
299 GCGTGGCGGATTTGGCGGCGATATATTTGTGAGCGTGGCACATAC 348
::: |||::: ||| ||| ||| ||| |||
34 rGThrLeuThrVallLeAspProGlnTYrAlaValAlaSerValLysVal 50
349 GCGCGC.....TATAACAATGTTGATTTGGTGGCGGAGG 383
|||
51 LysGlyAspGlnlLeSerTYrTYrGlyHSHsAsnGlyHSHsLeuAspVa 67
384 AAGCATCCCGATCGACACCGCTTTCTTACCAATTTGAAAAAATA 433
||||| |||::: ||| |||::: |||
67 lSerAsnAspGlnAsn.....GluTYrArgSerValAlaGlnAsnA 81

```


Ratio: 2.426 Gaps: 18
Percent Similarity: 63.472 Percent Identity: 37.794

Alignment block:

US-09-303-518D-653 x Q51168 ..

Align seg 1/1 to: Q51168 from: 1 to: 508

```

202 GCGAAGATATTGAGTTTACACAAAAAGGAGTGTGCGCAATC 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 AlatrhraspleuserValylsasnlysglnlysnlnleglyasn1 17
252 GATGACGAAGCCCGATGATGATTTTCTGTGTATCGGTATAC...G 298
    : : : : : : : : : : : : : : : : : : : : : :
17 aleuserasnValPromellleasPheSerValAlaspyAlasnlYsa 34
299 GCGTGGCGGCATTTGGCGGCGATCAATATATTGAGCGGTGCACAT... 345
    : : : : : : : : : : : : : : : : : : : : : :
34 rglleAlatrhrValAlasProglntYrAlalaserValylshisAla 50
346 .....AACGGCGCTTAACATGTTGATTT 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 LysAlagluValHisThrPheTyrTglnTyrAsnGlyHisAsnAs 67
372 TCGTGGCGGAGGAGCAATCCCGATCAGACGCCCTTTCTTACCAATG 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 pValAlasplysglnAsn.....GluTyrArgValY 78
422 TGAAGAAGAAATATTAAAGCGAGACTAACGCCATCTTATGCGGCG 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78 algluglnAsnAsnTyrGln.....ProHisYsAla 88
472 .....GATTATCATATGCCCGCTTT 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 TrpSerAlaSerAsnleuGlyArgleuGlnAspTyrAsnMetAlaArgph 105
492 GCACAAATTTGTACAGATGCAGAACCTGTGAGATGACAGTTATATG 541
    : : : : : : : : : : : : : : : : : : : : : :
105 eAsnlyspheValThrGluValAlaProthrAspAlaglyG 122
542 ATGGGTGG...AAATACGCTGATTAATAAATACCGTGTGTTGCA 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 lYcIlyleuAspThrTyrLysAspLysAsnArgPheSerSerPheValArg 138
589 ATGCGGCGAGCGACAGCAA.....TATTGCGGCTCGATGAGAGCAACC 632
    : : : : : : : : : : : : : : : : : : : : : :
139 ValGlyAlaGlyArgleuValTyrGlnYsglyAlaTyrHisGlnel 155
633 CATTAACCGCGAAGTTTCATAT.....CATATTGCAAGCGCATATT 673
    : : : : : : : : : : : : : : : : : : : : : :
155 uGlyAsn...GluYsglyTyrAspLeuArgAspleuserGlnAlaTyrA 171
674 CTGGCTCTCGGTGGC..... 690
    : : : : : : : : : : : : : : : : : : : : : :
171 rGlyrAlaAlaAlaGlyThrProTyrLysAspIleAsnIleAspGlnThr 187
691 ...AATACCTTTGACAAAATGATGAGTGGGACAGTCACTAAT 737
    : : : : : : : : : : : : : : : : : : : : : :
204 rAlaGlnGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrA 221
738 TACGAAAAAATTAACATATG...CCATATGATTTTATACCAACAGAG 784
    : : : : : : : : : : : : : : : : : : : : : :
204 rAlaGlnGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrA 221
785 GCTCATTTGGCGACAGTGGCTCCCAATGTTATCTATGATGCCCAAG 834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221 lYvalleuGlyAspSerGlySerProleuPheAlaPheAspLysGlnLys 237
835 CAAGAAGGTTAATTAATGAGGTATGCAACAGCAACCCCATATATAG 884
    : : : : : : : : : : : : : : : : : : : : : :
238 AsnGlnThrAlaPheleuGlyThrTyrAspTyrTTPAlaGlyTyrGly 254
885 AAAAGCAATGCTTCCAGCTAGTGTGTAAGATGTTGTATGATGAAGA 934

```

```

254 slyssentPrpGlnGluTyrAsnIleTyrLysLysGlnPheAlaAspLysI 271
    ||||| : : : ||| ||| ||| |||
935 TCCTTGTGTGAGATACCATTCCAGTATTTACGACCACTCAAAATGGG 984
    ||| |||
271 leYsglnArgAsp..... 275
985 AAATACTTTTAAACCAATTAATATGCGGCAAGAAATATGATGCCAA 1034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 .....AsnAlaGlyThrIleYsgLysAsnGlyG 285
1035 ACATAAACACTATCTCTACCTTATGATTAATAACAGCAACCGTTCAAT 1084
    : : : : : : : : : : : : : : : : : : : : : :
285 uHis...HisTyrAsnIleThrPheGlyThr..... 294
1085 TGTTTAATGTTTCTTATTCGACAGCAGCAAGAACCTGTTTATCATGCT 1134
    ||| : : : : : ||| ||| ||| ||| ||| ||| |||
295 ....AsnSerHisIleGlySerThrAla.....ValArgLeu 305
1135 GCAGTGGGGTCACAGTATATGACACCACTGAATATGAGAAATAT 1184
    ||||| : : : : : ||| ||| ||| ||| ||| ||| |||
306 AlaGlyAsnGlnIlyAsp.....AlaAsnAsnGlyGlnAsnVa 318
1185 TTCCTTTATGACAAAGAAAGTGAATTGATTAATACAGACATCA 1234
    : : : ||| : : ||| ||| ||| ||| ||| ||| ||| |||
318 lThrPheGluAsp.....AsnGlyThrLeuValleuAspGlnAsnIleA 333
1235 ACCAAGCGCGGCGGCTTTGATTTTATGAGGTATATTACGTCGTCGCT 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 snGlnGlyAlaGlyGlyLeuPhePheYsglyAspTyrThrValYsgly 349
1285 AAAACACAGCAA...ACGTGGCAGCGCGGCTTTCATATCATGATG 1331
    ||||| : : : ||| ||| ||| ||| ||| ||| ||| ||| |||
350 AlaAsnAsnAspIleThrTrpLeuGlyAlaGlyLysAspAlaAspGl 366
1332 CAGTACCGTTACTTGAAGATTAACGGCGTGGCAACGACCGCTGTCCA 1381
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 lYslyValValIleTrpGlnValLysAsnProAsnGlyAspArgLeuAla 383
1382 AAATGCGCAAGCAACGCTGTGTCAAGCAAGGCAAGCAAGCAAGC 1431
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 yslleGlyLysGlyThrleuGlnleuAsnGlyThrGlyValAsnGlnGly 399
1432 TCGGTCAAGCGTGGCGGAGGTAAGTCACTTATGATCAGAGCGGACGA 1481
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 GlnleuYsValGlyAspGlyThrValIleleuAsnGlnLysAlaAspse 416
1482 TCAAGCAAAAACAGAGCTTTAGTGAATCGCTTGTGTACGGCGACGG 1531
    : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 rAsnGlnLysValGlnAlaPheSerGlnValGlyIleValSerGlyArgA 433
1532 GGAGGTGCACTGATATGCGATATACGATTAACCGCGCAAACTGAT 1581
    ||||| : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 rGlyThrLeuValleuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyr 449
1582 TTCGGCTTTCGCGCGGAGCGTGTGATTTGAACGCGCATTCGCTTGGT 1631
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 PheGlyPheArgGlyGlyArgLeuAspAlaAsnGlnLysAsnAspLeuThrph 466
1632 CCACCGCATTAATAATACCGATGAAGGGCGATGTTGTCAACCAATC 1681
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
466 eGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHisAsnT 483
1682 AAGACAAAGAAATCCACGTTACCACTTACAGCAATTAAGATATTACTGA 1731
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 hrglyHisAlaSerThrIleThrleuThrGlyLysSerleuIleThrAla 499
1732 ACCGCGCAAT 1740
    |||
500 ProGlnAsn 502
seq_name: sp_bacteria:Q51164

```

seq_documentation_block:
 ID O51164 PRELIMINARY: PRT: 508 AA.
 AC O51164;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE IGA1 PROTEASE (FRAGMENT).
 GN IGA.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SM1166;
 RX MEDLINE=95302961; PubMed=7783620;
 RA Lombolt H., Poulsen K., Mogens K.;
 RT "Comparative characterization of the iga gene encoding Iga1 protease
 in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
 influenzae".
 RL Mol. Microbiol. 15:495-506(1995).
 DR EMBL; X82486; CA57869.1; -;
 DR MEROPS; S06.001; -;
 DR InterPro; IPR000710; IGA_S6.
 DR Pfam; PF02395; IGA1; 1.
 KW Protease.
 FT NON_TER 1
 FT NON_TER 508
 SEQUENCE 508 AA; 56025 MW; FFAF91F9BE01B38F CRC64;

alignment_scores:
 Quality: 845.50 Length: 549
 Ratio: 2.416 Gaps: 18
 Percent Similarity: 63.752 Percent Identity: 37.523

alignment_block:

US-09-303-518D-653 x O51164

Align seg 1/1 to: O51164 from: 1 to: 508

```

202 GCGAAGATATTGAGTTTACAAACAAAAGGAGGAGTGTGCGCAATC 251
    ||| |||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
1  A|A|ThrAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAsnAl 17
252 GATGACGAAGCCCGGATGATTTGTTGTGTATCGGTAAC...G 298
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
17  A|e|u|e|r|s|n|V|A|P|r|o|t|e|i|n|L|e|a|s|p|H|e|s|e|r|V|A|L|A|s|p|A|l|s|n|L|y|s|A 34
299 GCGTGGCGGATTTGGCGGCGCATATATTGAGCGGTGCACAT... 345
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
34  r|G|l|e|a|l|e|Th|r|V|a|l|V|a|l|s|p|r|o|t|e|i|n|r|a|l|a|V|e|r|V|a|l|L|y|s|H|I|s|A| 50
346 .....AACGGCGCTATTAACAATGTTGATT 371
51  L|y|s|a|l|a|G|l|u|a|l|H|I|s|Th|r|P|he|T|y|r|T|y|c|l|y|G|l|n|I|y|r|a|s|n|G|l|y|H|I|s|a|n|a|s 67
372 TGGTGGCGGAGGAGCAATCCCGATCAGACCGCTTTTCTTACCAATG 421
    |||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
67  p|v|a|l|A|l|a|s|p|L|y|s|G|l|u|a|s|n|.....G|l|u|y|r|a|r|y|V|a|l|y 78
422 TGAAGAATATTAATTAAGCAGGAGCTAGCGCATCCCTTATGGGGGC 471
    |||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
78  a|l|G|l|u|G|l|a|s|n|H|s|n|T|y|r|G|l|u|.....P|r|o|H|I|s|L|y|s|A| 88
472 .....GATTATCATATGCCGCTTT 491
89  T|I|p|e|r|a|l|a|s|e|r|a|s|n|L|e|u|G|l|y|A|r|G|l|u|G|l|u|a|s|p|L|y|r|a|s|n|e|t|a|l|a|A|r|G|p|H 105
492 GCACAATTTGTACAGATGACAGAACTGTGAGATACCAAGTTATATGG 541
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
105  e|A|s|n|L|y|s|e|V|a|l|Th|r|G|l|u|a|l|a|P|r|o|t|e|a|l|a|P|r|o|t|H|r|a|s|p|A|l|a|l|y|G 122
542 ATGGGTGG...AATACGCTGATTAAATAATACCTGATCGTTCGA 588
```

```

122  L|y|G|l|y|L|e|u|a|s|p|H|r|T|y|r|L|y|s|a|s|p|L|y|s|a|n|A|r|G|p|H|e|s|e|r|P|h|e|V|a|l|a|r|G 138
569  A|T|G|G|A|G|A|G|C|A|G|A|C|A|A|.....T|A|T|T|G|G|G|G|T|C|T|G|A|T|G|A|G|A|G|A|C|C 632
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
139  V|a|l|G|l|y|A|l|a|G|l|y|A|r|G|l|n|L|e|u|V|a|l|T|y|r|G|l|u|L|y|s|G|l|a|T|y|r|H|I|s|G|l|n|G|l 155
633  C|A|A|T|A|C|C|G|C|G|A|A|G|T|C|A|T|A|T|.....C|A|T|A|T|G|C|A|A|G|C|C|A|T|A|T|T 673
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
155  u|G|l|y|A|s|n|.....G|l|u|y|s|G|l|y|T|y|r|a|s|p|L|e|u|A|r|G|a|s|p|L|e|u|S|e|r|I|n|a|l|a|T|y|r|A 171
674  C|T|T|G|C|T|C|T|G|C|G|T|G|C|..... 690
171  r|G|l|y|r|a|l|a|l|e|a|l|G|l|y|Th|r|P|r|o|t|y|r|L|y|s|a|s|p|L|e|a|n|I|e|a|s|G|l|n|Th|r 187
691  ..A|A|T|A|C|T|T|G|C|A|C|A|A|A|T|G|A|T|C|A|G|T|G|G|G|G|C|A|G|T|C|A|C|T|T|A|G|G 737
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
188  M|e|t|a|s|n|Th|r|G|l|u|G|l|y|L|e|u|I|l|e|G|l|y|P|h|e|G|l|y|A|s|n|H|I|s|n|Th|r|H|I|s|T|y|r|S|e 204
738  T|A|G|G|A|A|A|A|A|T|T|A|A|C|A|T|A|G|C|.....C|C|A|T|A|T|G|C|T|T|T|T|T|A|C|A|A|G|A|G| 784
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
204  r|A|l|a|G|l|u|G|l|u|e|u|y|s|G|l|n|a|l|e|u|S|e|r|G|l|n|a|s|p|A|l|a|l|e|u|Th|r|a|s|n|T|y|r|G 221
785  G|C|T|A|T|T|T|G|G|C|A|G|A|G|T|G|C|T|C|A|C|C|A|A|T|G|T|T|A|T|C|T|A|T|G|A|T|G|C|C|A|A|A|G 834
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
221  L|y|V|a|l|e|u|G|l|y|A|s|p|e|r|C|l|y|S|e|r|P|r|o|l|e|u|P|h|e|a|r|P|h|e|a|s|p|L|y|s|G|l|n|L|y|s 237
835  C|A|A|A|G|T|G|T|A|T|T|A|T|G|G|G|T|A|T|T|G|C|A|A|A|G|C|A|C|C|T|A|T|A|T|A|G|G 884
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
238  A|s|n|G|l|n|Th|r|P|a|l|P|h|e|u|G|l|y|Th|r|T|y|r|a|s|p|L|y|r|P|a|l|G|l|y|r|G|l|y 254
885  A|A|A|G|C|A|A|T|G|G|C|T|C|A|G|C|T|A|G|T|G|T|A|A|G|A|T|G|G|T|C|T|A|G|A|A| 934
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
254  S|l|y|S|e|r|T|y|r|G|l|n|G|l|u|T|r|P|a|s|n|I|l|e|T|r|L|y|s|G|l|u|P|h|e|a|l|a|s|p|L|y|I 271
935  T|C|T|T|G|C|T|G|A|G|A|T|A|C|C|A|T|T|C|A|G|A|T|T|T|A|C|A|A|C|C|A|C|A|A|T|A|T|G|G 984
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
271  L|e|y|s|G|l|n|A|r|G|a|s|p|..... 275
985  A|A|T|A|C|T|T|T|T|T|A|C|G|A|C|A|T|A|T|A|T|G|G|G|C|A|G|A|A|A|T|G|C|A|T|C|C|A 1034
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
276  .....A|s|n|a|l|a|G|l|y|Th|r|I|l|e|y|s|G|l|y|A|s|n|G|l|y 285
1035  A|C|A|T|A|A|C|A|C|T|A|T|T|C|T|A|C|C|T|T|A|G|A|T|T|A|A|A|A|C|A|C|A|C|C|T|G|C|A|T 1084
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
285  u|H|I|s|.....H|I|s|T|r|a|s|n|I|e|Th|r|P|h|e|G|l|y|Th|r|..... 294
1085  T|G|T|T|A|T|G|T|T|C|T|T|A|T|C|C|G|A|G|A|C|A|G|A|C|A|C|A|C|C|T|G|T|T|A|T|C|A|T|G|C|T 1134
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
295  ....A|s|n|S|e|r|H|I|s|I|l|e|G|l|y|S|e|r|Th|r|A|.....V|a|l|A|r|G|l|e|u 305
1135  G|C|A|G|T|G|G|G|T|C|A|C|A|G|T|A|T|G|C|A|C|C|A|C|A|C|T|G|A|T|A|T|G|G|A|A|A|T|A|T 1184
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
306  A|l|e|G|l|y|A|s|n|G|l|u|a|r|G|a|s|p|.....A|l|a|s|n|A|s|G|l|y|G|l|n|a|s|n|V|a 318
1185  T|T|C|T|T|A|T|T|G|A|C|A|A|G|A|A|G|T|G|A|T|T|G|A|T|T|A|C|C|A|C|A|C|A|C|A|T|C|A 1234
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
318  I|Th|r|P|h|e|G|l|u|a|s|p|.....A|s|n|G|l|y|Th|r|e|u|V|a|l|l|e|u|a|s|p|G|l|n|a|s|n|I|l|e|a 333
1235  A|C|C|A|A|G|C|C|G|G|G|G|G|T|T|G|T|A|T|T|G|A|G|G|T|A|T|T|T|A|C|G|T|C|G|C|T 1284
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
333  s|n|G|l|n|G|l|y|A|l|a|G|l|y|L|e|u|P|h|e|P|h|e|L|y|s|G|l|y|A|s|p|L|y|r|Th|r|V|a|l|L|y|s|G|l|y 349
1285  A|A|A|A|C|A|C|G|A|A|.....A|C|G|T|G|C|A|A|G|C|G|G|G|G|C|T|C|A|T|A|T|G|A|G|T|G|A|G 1331
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
350  A|l|a|s|n|A|s|n|a|s|p|L|e|Th|r|P|r|e|u|G|l|y|A|l|a|G|l|y|I|l|e|a|s|p|V|a|l|A|a|s|p|G|l 366
1332  C|A|G|T|A|C|G|T|T|A|C|T|T|G|A|A|G|T|A|A|G|C|G|T|G|C|A|A|G|C|C|C|T|G|T|C|A 1381
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
366  L|y|L|y|s|V|a|l|V|a|l|T|r|G|l|n|V|a|l|L|y|A|s|n|P|r|o|k|s|n|G|l|s|p|A|r|G|l|e|u|a|l|a|l 383
1382  A|A|T|G|G|G|C|A|A|G|C|A|C|G|C|T|G|G|T|T|C|A|A|G|C|C|A|A|G|G|G|A|A|C|C|A|G|C 1431
    :|||:||||| ||| |||:||||| ||| |||:||||| ||| |||:|||||
```

```

383 ysllleGIySGlYthrlEuGlulIeasnglyThrclyValasngIngly 399
1432 TCAGTACGCGTGGCGAGGTAAGTCATCTAGATCAGCAGCGGAGCA 1481
400 GlulEuYsValIyAspGlyThrValIleuAsngInlyAspse 416
1482 TCAGGCAAAAACAGCCTTAGTGAATCGCGTGTGACGGCGAGG 1531
416 rAsngInlyValSerAlaPheserInlyAlglyIleValArgGlyArg 433
1532 GAGCGGTGCACTGAATGCCGATATCAGTTCACCCGCAAACTCTAT 1581
433 lyThrlEuValIleuAsnSerProAspGlnIleasnProAsnInly 449
1582 TTGGGCTTGGCGGAGCGTGTGATTTGAACGGCATTCGCTTCGTT 1631
450 PhedIyPhenArgIyGlyArgLeuAspAlaAsnglyAsnspReuThr 466
1632 CCACCGCATTCAAATATACGATGAGGCGGATGATTCACACGATTC 1681
466 eGlulHsIleArgAsnValAspGlnlyAlaArgIleValAsnHsAsn 483
1682 AAGACAAAGATCCACCGCTTACCATTCAGCGCAATAAAGATTTACT 1728
483 hrAspArgAlaSerThrlIethrlEuThrclySserLeuIethr 498
seq_name: sp_bacteria:Q51173

```

```

seq_documentation_block:
ID Q51173 PRELIMINARY; PRT; 507 AA.
AC Q51173;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IGAL PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG093;
RX MEDLINE=95302961; Pubmed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding IGAL protease
in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
influenzae."
RL MOL. Microbiol. 15:495-506(1995).
DR EMBL: X82482; CAA57865.1;
DR MEROPS: S06.001;
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGAL; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 507
SQ SEQUENCE 507 AA; 55956 MW; 80ETBA3F04EBBA82 CRC64;

```

```

alignment_scores:
Quality: 840.00 Length: 546
Ratio: 2.421 Gaps: 17
Percent Similarity: 63.553 Percent Identity: 37.912

```

alignment_block:

US-09-303-518D-653 x Q51173 ..

Align seg 1/1 to: Q51173 from: 1 to: 507

```

202 GCGAAGATATTGAGTTTACACAAAAGGAGGATTTGGCGCAATC 251
|||||
1 AlathrAspLeuSerValIyAsnlySngInlyAsnIleGIyAsnAl 17
252 GATGACGAAGCCCGATGATTTTCTGTGTGTAATCGCGTAAC...G 298

```

```

17 ..... 34
17 aleuSerAsnValProMetIleAspPheSerValAlaAspValAsnlySa 34
299 GCGTGGCGGCAATTTGGCGGCGATCAATATATTGAGCGTGCGACAT... 345
34 rglleAlaThrValValAspProGlnlyAlaValSerValIyAsnHsAla 50
346 ..... 371
51 lysAlaGluValAlHsThrPheTyTrGlyGlnlyrAsngInlyHsAsnAs 67
372 TGGTGGGAGGAGGAGCAATCCCGATCAGCACCCTTTTCTTACCAAAATG 421
67 pValAlaAspIySngIuAsn.....GlnTyTrArgValV 78
422 TGAAGAATATATATTAAGACGAGGACTAAGCGGCATCTTATGGCGGC 471
78 alGluGlnAsnInlyrGluPro.....HslyAlaAlaTrpIly 90
472 ..... 497
91 AlaserAsnleuGlyArgleuGlnAspTyTrAsnMetAlaArgPheAsnly 107
498 ATTTGTACAGATGCGAAGACCTGTGAGATGACAGGTATATGATGGGT 547
107 spheValThrGluValAlaProIleAlaProThrAspAlaGlyIlyL 124
548 GG...AAATACGCTGATTTAAATAATACCGTACGTCGTTGATCGA 594
124 euAspThrTyTrIyAspIyAspIyAsnAlArgPheSerPheValAlaArgVal 140
595 GCAGCGAGCAAA...TATTGGCGGTGTGATGAAGCAACCAATATACG 641
141 AlagIyArgGlnleuValTyTrGlnlySgIyAlaTyTrHsGlnGlnly 157
642 CGAAGTTCAATAT.....CATATGCAAGCGCATATTTCTGGCTCG 682
157 sGlnlySgIyTyTrAspLeuArgAspLeuSerGlnAlaTyTrArgTyAla 174
683 TCGGTGCG.....AATACC 696
174 IeAlaIyThrProTyTrIyAspIleAsnIleAspGlnThrMetAsnThr 190
697 TTTCGCAAAATGATCAGGTGGCGACAGTCACCTTAGTAGGAGAAA 746
191 GluGlyLeuIleGlyPheGlyAsnHsAsnlySngInlySerAlaGln 207
747 AATTAACATAC...CCATATGCTTTTACCACAGAGGCTCATTTG 793
207 uLeuTySngInAlaIeuserGlnAspAlaLeuThrAsnTyTrGlyValleuG 224
794 GCGACAGTGGCTCACCAATGTTATCTATGATGCCAAAGCAAAAGTGG 843
224 IyAspSerIySerProLeuPheAlaPheAspIySngInlySngInlyrP 240
844 TTAATTATGGGATTTGCAACAGGCAACCCCTATATAGAAAAGCAA 893
241 ValPheleuGlyThrTyAspTyTrAlaGlyTyGlyLylySserTr 257
894 TGGCTTCACAGTACTGTAAGATTTGTTCTATGATGAATCTTTGCG 943
257 pGlnGlnTyTrAsnIleTyTrIySngIySngIyPheAlaAspGlnIle 271
944 GAGATACCCATTCATTTCTACGACACATCAACAAATGGAATACTTT 993
271 ..... 271
994 TTTAACGACATATATATGCGCAGGAAAATCGATGCCAAACATTAACA 1043
272 .....LysGlnArgAspAsnAlaGlyThrIleTySngIyTyGlyGln 286
1044 CTATTCCTCACTTATATAGATTAAACACGAACCGTTCATTTGTTAAG 1093
|||||

```

```

286 shis.....TrpLysThrThrGlyThr.....AsnS 295
1094 TTCTTTATCCGACAGCAAGAACCTGTTATCATCTGCGAGGCGG 1143
      ::::::::::::::|
295 erhisileglySerThr.....AlaValAlaGlu 304
      ::::::::::::::|
1144 GTCACAGCTATTCAGCCAGACTGAAATATGAGAAAATATTCCTTAT 1193
      ::::::::::::::|
305 AlaasnngluArgasp...AlaasnnglyGluasnvalThrhe.. 319
      ::::::::::::::|
1194 TGCACAAAGAAAGTCATGATGATCTACTACACACATCACCAAGCG 1243
      ::::::::::::::|
320 ....GluasnnglyThrLeuValLeuaspGluasnileasnnglyYA 335
      ::::::::::::::|
1244 CGGCGGCTTGTATTTGAGGTAATTTTCGCTCGCTTAACAAACAC 1293
      ::::::::::::::|
335 laglyGlyLeuPhePhePheGlyAspPyrThrVallysglyAlaasn 351
      ::::::::::::::|
1294 GAA...ACGTGGCAAGCGCGCGGCTTCATATCATGATGAGGACGT 1340
      ::::::::::::::|
352 AspIleThrTrpLeuGlyAlaGlyIleaspValAlaaspolyLysys 368
      ::::::::::::::|
1341 TACTGGAAAGTAACGGCGGTGGCAAGCCGCTGTCCAAATGCGCA 1390
      ::::::::::::::|
368 lvalitPglInvalLysasnProasnnglyAspArgLeuAlaLysile 385
      ::::::::::::::|
1391 AAGCGACGCTGCTGTTCACAGCCCAAGGGAACCAAGCGCTGCGCAG 1440
      ::::::::::::::|
385 ysglyThrLeuGluIleasnnglyThrGlyValasnnglyGluLeu 401
      ::::::::::::::|
1441 GTGGGGGACGGTAAAGTCATCTAGATCAGCAGCGGACGATCAGGCA 1490
      ::::::::::::::|
402 ValGlyAspPolyThrValIleLeuasnnglyAlaaspAlaaspLys 418
      ::::::::::::::|
1491 AAACACAGCCTTGTAGTAATCGCTTGGTCAGCGGAGGAGCGGTCG 1540
      ::::::::::::::|
418 svalGlnAlaPheSerGlnValGlyIleValSerGlyArgAlaGly 435
      ::::::::::::::|
1541 AACTGAATGCCATTAATCAAGTTCACCCGACAACTATTTCCGGCTT 1590
      ::::::::::::::|
435 alLeuasnserSerasnnglnIleasnProaspasnLeuThrPheGly 451
      ::::::::::::::|
1591 CGGCGGAGCGTTGGATTTGAAACGGGCATTCGCTTCTCCACCGCAT 1640
      ::::::::::::::|
452 ArgGlyValArgLeuaspAlaasnnglyAsnaspLeuThrPheGlu 468
      ::::::::::::::|
1641 TCAAAATACCGATGAAGGCGCATGTGTCAACCAATCAAGCAAG 1690
      ::::::::::::::|
468 earGAsnValaspGluGlyAlaArgIleValasnHisasnThrGly 485
      ::::::::::::::|
1691 AATCCACCGTTACATTCAGCGCAATAAAGATTTACT 1728
      ::::::::::::::|
485 laseThrIleThrLeuThrnglyLysSerLeuIleThr 497
      ::::::::::::::|
seq_name: sp_bacteria:069746
seq_documentation_block:
ID 069746 PRELIMINARY; PRT; 192 AA.
AC 069746;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ICA PROTEASE (FRAGMENT).
OS Neisseria lactamica.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=486;
RN [1]
RP SEQUENCE FROM N.A.
RC SPMRAIN=NL3327;
RA Sayers J.R., Vitowski S., Read R.;
RT "Strains of Neisseria lactamica carry an Iga1 protease gene homologue
and some express Iga1 protease activity.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL: AJ001740: CAN04965.1; -.
KW MEROPS, S06.001; -.
FT NON_TER 1 1
FT NON_TER 192 192
FT NON_TER 192 192
SQ SEQUENCE 192 AA; 21385 MW; E4C6FE902FFA08F8 CRC64;

alignment_scores:
  Quality: 732.50
  Ratio: 4.334
  Percent Similarity: 84.500
  Percent Identity: 72.000

alignment_block:
US-09-303-518d-653 x 069746 ..

Align seg 1/1 to: 069746 from: 1 to: 192

184 GGCAGATTGTCAGTCGCGGCGCAAGATATGAGGTTTACACAAAAGG 233
      ::::::::::::::|
1 GilyuSpheAlaValAlaGlyAlaLysAspIleGluValIlyrAsnLysG 17
      ::::::::::::::|
234 GGAGTTGGTCGCAAAATGATGACGAAACCCCGATGATGATTTCTG 283
      ::::::::::::::|
17 GilyuValAlaGlyLysSerMetThrLysAlaProMetIleAspPheSer 34
      ::::::::::::::|
284 TGGATTCGCTTAACGCGCTGGCGGCTTGGCGGCGATCATATATTGTG 333
      ::::::::::::::|
34 alvalSerArgasnnglyAlaAlaLeuValGlyAspGlnIlyrIleVal 50
      ::::::::::::::|
334 AGCGTGCGCATTAACGCGGCTATTAACAATGTTGATTTGGTCGCGAGG 383
      ::::::::::::::|
51 SerValAlaHisasnnglyGlyTrAsnAsnValAspPheGlyAlaGlu 67
      ::::::::::::::|
67 ySerAsnProaspGlnHisArgPheSerTyGlnIleValLysArgSna 84
      ::::::::::::::|
434 ATTATAAGCAGGAGCTAACGCGCATTCCTTATGGCGGCGGTTATTCATATG 483
      ::::::::::::::|
84 snTyLysPro..AspAsnserHisProTyTrAsnnglyAspTyTrHisMet 99
      ::::::::::::::|
484 CGGCGTTTGCACAAATTTGTCACAGATGCAAGACCTGTGAGATGACAG 533
      ::::::::::::::|
100 ProArgLeuHisLysPheValThrAspAlaGluProValGluMetThr 116
      ::::::::::::::|
534 TTAATGATGATGGCGGAATATGCTGATTTAAATATACCTGATGCTG 583
      ::::::::::::::|
116 yAspMetArgGlyAsnThrTySerAspLysGluLysTyTrProGluArg 133
      ::::::::::::::|
584 TTGCATATCGAGCAGCGACAGCAATATGCGCGCTGTGATGAAGCAAGCC 633
      ::::::::::::::|
133 alArgIleGlySerGlyHisHisIlyrTrpArgTyTrAspAspLys... 148
      ::::::::::::::|
634 AATGACCGCGAAAGTTCATATCATATTCGAAGCGCATATTTCT...TG 677
      ::::::::::::::|
149 .....HisGlyAspLeuSerTyTrSerGlyAlaArg 158
      ::::::::::::::|
678 GCTGTCGGTGGCAATACCTTTGCACAAAATGATGATGAGTGGGACAG 727
      ::::::::::::::|
158 PheIleIleGlyAsnThrHisMetGlnGlyTrpGlyAsnAsnnglyVal 175
      ::::::::::::::|
728 TCAACTAGCTAGAGCAAAATTAACATATGCCCATATGCTTTTATACCA 777
      ::::::::::::::|
175 alSerLeuSerGlyAspValArgHisAlaasnspTyTrGlyProMetPro 191
      ::::::::::::::|
seq_name: sp_bacteria:069747
seq_documentation_block:
ID 069747 PRELIMINARY; PRT; 194 AA.
AC 069747;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)

```



```

34 ysaarglylethLeuValAspProGlnTyrlleValSerValIshs 50
346 AAC.....GGCGCTATACAAAGTTGA 368
51 AlaHisIshsIleAsnAspPheTyrlleHisIshs..... 64
369 TTTGGTGGGAGGAGCAATCCGATCCAGCGCTTTCTTACAA 418
65 ...GlyHisArgAspValSerAspAspIshs.....Tyrlle 78
419 TTGTGAAGAATAATTAAGCAGGAGCTACGGCAT.....CCT 462
78 alValIshsIshsValIshsValIshsValIshsValIshs 94
463 TATGGCGGATATCATATGCCGGTTCACAAATTTGTACAGATC 512
95 ArgLeuAspAspIshsIshsIshsIshsIshsIshsIshs 111
513 AGAACCTGTGATGACCAATTATATGATGGTGAA...TAGCGT 559
111 lAlaProIshsIshsIshsIshsIshsIshsIshsIshs 128
560 ATTAAATTAATACCTGATCGTTGCAATGCAAGCAGCAGCAATAT 609
128 sPylsIshsIshsIshsIshsIshsIshsIshsIshsIshs 144
610 TGGCGCTCGAT.....GAGCAGCAACCAATACCGCA 644
145 ValTyrlleIshsIshsIshsIshsIshsIshsIshsIshs 161
645 AAGTATATCATATGCAAGCAGCATATCTGGTCCGCGGCAAA 694
161 lIleIshsIshsIshsIshsIshsIshsIshsIshsIshs 177
695 CCTTTCAC.....CAAAATGATCAGGTGCGCAGATC 729
177 rOtyrlleIshsIshsIshsIshsIshsIshsIshsIshs 193
730 AACTTGGTAGCAAAATTAACAT..... 756
194 GlyPheIshsIshsIshsIshsIshsIshsIshsIshsIshs 210
757 .....AGCCATATGTTTACCAACAGAGCTCAT 790
210 uSerGlnAspProLeuIshsIshsIshsIshsIshsIshs 221
791 TTGGCAGATGCTCACCACCATTTATCTATGATCCCAAAAG 840
222 ...GlyAspSerGlySerProLeuPheAlaPheAspIshsIshs 237
841 TGCTTATATATGGGTATTCACAAACAGCAACCCAT..... 879
238 TrpValPheIshsIshsIshsIshsIshsIshsIshsIshs 248
880 .....ATAGCAAAAGCAAT.....GGCTCCAGTAGTCTGTAAGATT 919
248 agIshsIshsIshsIshsIshsIshsIshsIshsIshsIshs 264
920 GGTTCATATGAATCTTTGTCGAGATCCCAATTCAGTATTCAGAA 969
265 ...PheAlaAspAsnIshsIshsIshsIshsIshsIshsIshs 278
970 CCACATCAAAATGGAATCTTTTACGACAAATTAATATGCGCAG 1019
279 ProPheSerIshsIshsIshsIshsIshsIshsIshsIshs 295
1020 AAAATGATGCAACATTAACATATCTACCTTATATATTAATAA 1069
295 rGluIshsIshsIshsIshsIshsIshsIshsIshsIshs 310
1070 CACGACCGTTCAATGTTTAAATCTTATCCGACAGCAAGAA 1119

```

```

310 erAspProAsnArgLeuValAsnPheGlnIshsIshsIshsIshs 322
1120 CCTTTATATCATGCTGACAGGTGGGCTCACAGTATATGACCAAGTAA 1169
323 .....HisIshsIshsIshsIshsIshsIshsIshsIshs 325
1170 TAAATGCAAAATATTTCTTATTTGAC.....AAGCAAAAGT 1210
325 nThrGlyGlnAsnValIshsPheAspAspSerIshsIshsIshsIshs 342
1211 AATGATATCTTACCAACATCATACCAAGCGCGGCTTTGATTT 1260
342 hrLeuIshsIshsIshsIshsIshsIshsIshsIshsIshsIshs 358
1261 GAGGTATATTTACGGTCTGCTCCCTAAACACAGCA...ACGTGCAAG 1307
359 LysIshsIshsIshsIshsIshsIshsIshsIshsIshsIshsIshs 375
1308 CGCGGCGCTTATCATATGATGACAGTACCGTTTACGAAATGAA 1357
375 yGlyGlyIshsIshsIshsIshsIshsIshsIshsIshsIshsIshs 392
1358 CGGTGCAACAGCAGCGCTGCTCAAAATGCAAGCAGCAGCTGCTGTT 1407
392 snProGlyIshsIshsIshsIshsIshsIshsIshsIshsIshsIshs 408
1408 CAGGCAAAAGGGAACCAAGCGCTGCTGCGGCGCAGGTAAGT 1457
409 GluGlyIshsIshsIshsIshsIshsIshsIshsIshsIshsIshs 425
1458 CATCTTATGATGACGAGCGGCGCAGATCAAGCAAAACCAACCTTATG 1507
425 lValLeuIshsIshsIshsIshsIshsIshsIshsIshsIshsIshs 441
1508 AATCGCTTGTGTACGCGGCGGCGGAGCGGTGCACACTGAATCCGATAAT 1557
441 erValIshsIshsIshsIshsIshsIshsIshsIshsIshsIshs 457
1558 CAGTTCACCCGCAACACTTATTTGCGTTCGCGGCGCA 1599
458 GluValAspProAsnSerIshsIshsIshsIshsIshsIshsIshs 471

```

seq_name: sp_bacteria:069745

```

seq_documentation_block:
ID 069745 PRELIMINARY; PRT; 190 AA.
AC 069745;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
OS Neisseria lactamica.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=486;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NL3293;
RA Sayers J.R., Vltowski S., Read R.;
RT "Strains of Neisseria lactamica carry an Iga1 protease gene homologue
RL and some express Iga1 protease activity."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ001738; CAA04963.1; -.
DR MEROPS; S06.001; -.
KW protease.
FT NON_TER 1
FT NON_TER 190
SQ SEQUENCE 190 AA; 21016 MW; 4A524869DB83BAF1 CRC64;

```

alignment_scores: 658.50 Length: 192
 Quality: 4.015 Gaps: 4
 Percent Similarity: 85.417 Percent Identity: 67.188

alignment_block:

US-09-303-518D-653 x 069745 ..

Align seg 1/1 to: 069745 from: 1 to: 190

```

205 AAGGATTTGAGTTTACAAACAAAAAGGAGCTTGGTGGCAATTCAT 254
1 LysAsp1IeGIvAlTyRAsnLysG1uLeuValG1LysSerMe 17
255 GAGGAAAGCCCGATGATGATTTCTGCGATGCGTAACGGCGTGG 304
17 tThrLysAlaProMeL1eAspSerValAlSerAlaGsnG1LysVal 34
305 CGGCATTGGCGGCGATCAATATATTTGTAGCGTGACACATACGCGC 354
34 laAlaLeuValG1LysAspG1nTyR1LeValSerValAlaH1snValG1 50
355 TATTAACAATGTTGATTTGGTGGCGGAGGACCAATCCCGATCAGCAGC 404
51 TyTAtGAspValAspPheG1LysAlaG1uG1LysAsnProAspG1nH1sAr 67
405 CTTTTCTTCAACAATTTGTAAAGAAATATATTAAAGCAGGACTAACG 454
67 gRheserTyR1Lys1LeAlaLysAlaGsnAsnTyR1Lys...AsnAspG1uT 83
455 GGCATCCTTATGGCGCGATATGATATGCGCGCTTTGACAAATTTGTC 504
83 hR1sPrG1uTyR1uLysAspTyR1H1sAsnProAlaG1uH1sLysPheVal 99
505 ACAGATGCGAACCTGTGATGATGACAGCTTTATGATGGTGGAAATA 554
100 ThG1uAlaAlaPro1LeAspMeThR1SerAspMetAspG1LysLysTy 116
555 CGCTGATTTAAATAAATACCTGATGCTGTGATGCGAGCAGCAGCAGC 604
116 rThRAspArgThR1LysTyRProG1uArgValAlaG1LeG1LysSerG1TyR 133
605 AATATTTGGCGGTGTATGATGAAGACCAATACCGGAAAGTTTCATAT 654
133 LnPheTyRArgAsnAspG1nAspLysG1LysP... 143
655 CATATTTGACAGCGCATATTTCTGCTCGTGGTGGCATACCTTTGCACA 704
144 G1nValAlaG1LysAlaTyR1H1sTyRLeuThR1AlaG1LysAsnThR1H1sAsnG1 160
705 AATGGATCAGGTGGTGGCAGACAGTCACTAGTACGCAAAAAATTAAC 754
160 nG1LysLysAlaG1LysG1TyRTPSerLeuSerG1LysP...ValArg 176
755 ATAGC...CCATATGCTTTTTCACA 777
176 IsAlaG1LysnTyR1G1LysPro1LePro 184

seq_name: sp_bacteria:068900

seq_documentation_block:
ID 068900 PRELIMINARY; PRT; 1295 AA.
AC 068900;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE PRT. PRECURSOR.
GN PRT.
OG Escherichia coli.
OC Plasmid pO42.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=042;
RX MEDLINE=98298033; PubMed=9632580;

```

```

RA Eslava C., Navarro-Garcia F., Czeckulin J.R., Henderson I.R.,
RA Cravioto A., Nataro J.P.;
RT "Prt, an autotransporter enterotoxin from enterogaegregative
RT Escherichia coli."
RL Infect. Immun. 66:3155-3163(1998).
DR EMBL; AF056581; AAC26634.1; -.
DR MEROPS; S06.004; -.
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGA1; 1.
KW signal; Plasmid.
FT SIGNAL. 52 POTENTIAL.
SQ SEQUENCE 1295 AA; 139768 MW; 9C6AEF7E345AD429 CRC64;

```

alignment_scores:

Quality: 657.50 Length: 1489
 Ratio: 0.933 Gaps: 65
 Percent Similarity: 47.347 Percent Identity: 22.364

alignment_block:

US-09-303-518D-653 x 068900 ..

Align seg 1/1 to: 068900 from: 1 to: 1295

```

160 TATGCGCACTTTGGCCGAATAATTAAGCCAGTTTGCATGCGGCGCAAGA 209
111 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65 TyRLeuAspLeuAlaG1nAsnLysG1LysAlaPheG1nProG1LysThR1H1 81
210 TATTGAGGTTTACACAAAAGGGAGTTGGTGGCAATTCATGACGA 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 sValLys1LeLysLeuLysAspG1LysP...ThAspPheSerPhePro 96
260 AAGCCCGCATGATTTTCTGCTGATGCGGTACAGCGCGTGGCGCA 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96 lAlaProValProAspPheSerSerAlaThR1AlaSnG1LysAlaAlaThR 112
310 TTGGCGGCGCATCATATATTGTGAGCGTGCGACATACGCGGCTATGA 359
||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 Ser1LeG1LysAlaTyRAlaValThRValAlaH1sAsnAla..... 126
360 CAATGTTGATTTTGGTGGAGGAAACCAATCCCGATCAGACCGCTTTT 409
127 .....LysAsnLysSerSerAla 133
410 CTTAACAATTTGTGAAGAATATATTAACGACGAGTCAAGCGCCAT 459
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 snTyRGlntThTyR1G1LysThRGlntTyR...ThR1n1LeAsnArgMet 148
460 CCTTATGCGCGCATATATATGCGCGCTTTGACAAATTTGTGCACGA 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 ThRThR1G1LysAsnAspPheSer1LeG1nArgLeuAsnLysTyRValAlaG1 165
510 TGCAGAACCTGTTGAGATGACCACTTATATGATGGTGGAAATACGCTG 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 uThRArgG1LysAlaAsp...ThRSerPhe.....AsnTyRAsnG 177
560 ATTTAAATATAA...TACCTGATCGT..... 582
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177 LuAsnAsnG1nAsn1Le1LeAspArgTyR1G1LysAlaSpValG1LysnG1 193
583 .....GTTGCAATGAGACAGCGAGA..... 603
194 LysLysG1u1Le1LeG1LysPheArgValG1LysSerG1LysnThRThR1Phe 210
604 .....CAATATTTGGCGGTCTGATGATGAAGACGAACCAATA 637
210 rG1Lys1LeLysThRSerG1nThTyRGlntAlaAsp.....LeuL 223
638 ACCGCAAGTTCAATATATATGCAAGCGCATATCTGCTGCTGCTGCT 687
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 euser1aSer1LeuPheH1s1LeThRAsnLeuArgAlaAsnThRValG1Lys 239
688 GGCATATCTTTGACAAAATGATCAGGTGGTGGCAGTCACTTAAAG 737

```



```

|||||
240 GlyAsn..... 241
738 TACGAAAAAATTAAACATAGCCCATATGCTTTTACCACAGAGGCT 787
|||||
242 .....LysValGluTyrGluAsnAspSerTyrPheThrAsnLeuThr 256
788 CATTGGCGACAGTGGCTACCAATGGTTATCATGATGCCCAAAAGCA 837
|||||
256 HisnGlyAspSerGlySerGlyValTyrValPheAspAsnLysGluAsp 272
838 AAGTGGTTATTAATGGGCTATTCGCAACAGCCACCCCTATATAGGAA 887
|||||
273 LysTrp..... 274
888 AAGCATGGCTTCACCTAGTCTGTAAGATGGTTCTATGATGAATCT 937
|||||
275 .....ValL 276
938 TTGCTGAGATACCATTCAGTATCTTACGACACATCAAAATGGGAA 987
|||||
276 euLeuGlyThrThrHisGlyIleIle..... 284
988 TACTTTTAAAGCAATATATATGCGGAGGAAAAATGATCCAAACA 1037
|||||
285 .....GlyAsnGlyLys.....ThrG 290
1038 TAAACACTATCTCTACCTTATAGATTAAACA.....CGAACGCTTC 1081
|||||
290 nLysThrTyrValThrProPheAspSerLysThrThrAsnGluLeuLysG 307
1082 AATGTGTT.....AATGTTCTTTATCCGACAGCAAGAGAACCTGTT 1125
|||||
307 InLeuPheIleGlnAsnValAsnIleAspAsnAsnThrAlaThr..... 321
1126 TATCATGCTGCAGTGGG.....GTCAACAGTTATCG 1157
|||||
322 .....IleGlyGlyLysIleThrIleGlyAsnThrThrGlnAspI 336
1158 ACCGAGCTGAATATGAGNAAATATTCCTTATTCACAAAGAAAG 1207
|||||
336 eGluLysAsnLysAsnAsnGlnAsnLysAspLeuValPheSerGlyGlyG 353
1208 GTGATTTGATTACTTACCAGCAACATCAACAGCGCGGGCTGTGTAT 1257
|||||
353 LysLysIleSerLeuLysGluAsnLeuAspLeuGlyTyrGlyGlyPheIle 369
1258 TTGAGGGTAAAT.....TTTACGGTCTGCCTTAA.....AACACGA 1295
|||||
370 PheAspLysAsnLysLysTyrThrValSerAlaGluGlyAsnAsnAsn 386
1296 AAGCTGCAAGCGCGGCTTCATATCACTGATGAGCAGTACCGTTACT 1345
|||||
386 IThrPheLysGlyAlaGlyIleAspIleGlyLysGlySerThrValAsp 403
1346 GAAAGTAAACGGCGTGGCAACAGCCGCTGTCCAAATGGCAAAAGC 1395
|||||
403 rPheAsnIleLysTyrAlaSerAsnAspAlaLeuHisLysIleGlyGly 419
1396 ACGGTG...CTGTTCAAGCCAAAGGGAACCAAGGCTCGGTACGCT 1442
|||||
420 SerLeuAsnValIleGlnAlaGlnAsnThrAsn.....LeuLysThr 433
1443 GGGGAGGTAAGTCAATCTTATGATCAGCGGAGCATCAAGCAAA 1492
|||||
433 rGlyAsnGlyThrValIleLeu.....GlyAlaG 443
1493 AACAGCGCTTAACTGAATCGCGCTGTGTCAGCGGAGGAGCGGTGCA 1542
|||||
443 InLysThrPheAsnAsnIleTyrValAlaGlyProGlyThrValGln 459
1543 CTGAATGCCATTAATCAGTTCACCCGCAAA.....CTCTATT 1583
|||||
460 LeuAsnAlaGluAsnAlaLeuGlyGlyLysAspTyrAlaGlyIlePhe 476
1584 CGGCTTCGGCGGAGGAGCTTTCGATTTGAACGGGCTTCGCTTC 1633
|||||
476 eThrGluAsnGlyGlyLysLeuAspLeuAsnGlyHisAsnIleThrPhe 493
1634 ACCCATTCAAAAATACCGATGAGGGCGGATGATTCACACCAATCAA 1683
|||||
493 yLysIleAlaIleAlaThrAspSerGlyThrThrIleThrAsnSerAsnThr 509
1684 GACAAAGATCCACCGTTACCTTACAGGCAATTAAGATATTACTACAAC 1733
|||||
510 ThrLysGluSerValLeuSerVal..... 517
1734 CGGCAATACAAACAATTGATACCAAAAGAAATTCCTACACGGTT 1783
|||||
518 ....AsnAsnGlnAsn.....AsnTyrIleTyrHisGly. 527
1784 GGTTCGGGAGAAAGATGCAACCAAAAGAGCGGCTTCATCTGAAT 1833
|||||
528 .....AsnValAspGlyAsnValArgLeuGluHis 538
1834 TACCAACCGGAGAGGATCGCATTACTGCTTCCGGCGGACAA 1883
|||||
539 LeuAspThrLysGlnAspAsnAlaArgLeuIleLeuAspGlyAspIleG 555
1884 TTTAAGCGCAATATCAGCAAAACAAACGCAAACTGTTTTCAGCGCA 1933
|||||
555 nAlaAsn...SerIleSerIleLysAsnAlaProLeuValMetGlnGlyH 571
1934 GACCGACACCGCAGCCTTAC.....AATCAT..... 1959
|||||
571 ISAlaThrAspHisAlaIlePheArgThrThrLysThrAsnAsnCysPro 587
1960 .....TTAGCAAGGGGTGTCACAAATGGAAGGTATCCACAAAG 2000
|||||
588 GluPheLeuGlyGlyAlaAspThrValThrArgIleLysAsnAlaGluAs 2000
2001 AGAAATCGTGGGACAC..... 2019
|||||
604 nSerValAsnGlnLysAsnLysThrThrTyrLysSerAsnAsnGlnValS 621
2020 .....GATTGATGACCGCACATTAAAGCGGAAC 2052
621 eAspLeuSerGlnProAspTrpGluThrArgLysPheArgPheAspAsn 637
2053 TTCATATTCAGGGCGGACAGCGGTGGTTCCCGCAATGTTCCGAACT 2102
|||||
638 LeuAsnIleGluAspSerSerLeuSerIleAlaIleAsn...AlaAspVa 653
2103 GGAAGCGCATTTGCAATTAAGCAATCAGCCCAACGACTTTTCGGTGTG 2152
|||||
653 IGlIleAsnIleGlnAlaLysAsn.....SerValIleAsnIleG 667
2153 CACCGCAT..... 2160
667 LysAspLysThrAlaTyrIleAspLeuTyrSerGlyLysAsnIleThrGly 683
2161 .....CAAGCCACACAAATCTGTACAG 2183
684 AlaGlyPheThrPheArgGlnAspIleLysSerGlyAspSerIleGlyG 700
2184 TTCGACTGAGCGGCTGTGACAAGTTGTAACGAAAAACCATTTAC...G 2230
|||||
700 uSerLysPheThrGlyGlyIleMetAlaThrAspGlySerIleSerIleG 717
2231 ACGATTAAGTGAATGCTTATTCATGACCAAGCCGACATCAGAGGCAATGTC 2280
|||||
717 LysAspLysAlaIleValThrLeuAsnThrValSerSerLeuAspArgThr 733
2281 AGCCTTGCCGATCAGCGCTATTAAATCAGACAGACTTGCACACACT... 2328
|||||
734 AlaLeuThrIleHisLysGlyAlaAsnValThrAlaSerSerLeuPhe 750

```

2329AACGCAATCTTAGTCAGCGAGACGACGCACTATACGGTTACGC 2374
 750 ethrThrSerAsnIleYserSerglyAsp 760
 2375 GCACGCGCACCAAAAGCGACCTACCGCTGCGCAATGCCAAGCA 2424
 761LeuThrLeuThrGlyAlaThrGluSer 769
 2425 ACA.....TTTAAATCAAGCCACATTTAAAGGCA 2453
 770 ThrGlyGluIleThrProSerMetPheTyrAlaAla.....GlyGlyTyr 784
 2454 CACATGGCTTGGACCAATGCTTTCATTTATATCAACCAACGCGGTAC 2503
 784 rGluLeuThrGluAspGlyAlaAsnPherThrAlaYsAsnGlnAlaSerV 801
 2504 AAAAGCGCACTGACGCTTTCGACACAGCTAAGCAACGTAAGCCAT 2553
 801 alThrGlyAspIleYserSerglyAlaAlaYsLeuSerPheGlySer 817
 2554 TCCGCACTCAACGCAATGTCTCCCTAGCCGATTAAGCAGATTTCCATTT 2603
 818AlaAspLysAspAsnSerAlaThr 825
 2604 TGAACACGCGCTTACCGGAAAAATCAGCGCGCAAGATACG... 2649
 825 rArgTyrSerGlnPheAlaLeuAlaMetLeuAspGlyPheAspThrSert 842
 2650GCATTACACTTAAAGACAC 2670
 842 yriGlnIySerIleYsAlaAlaGlnSerSerLeuAlaMetAsnAla 858
 2671 GAATGACGCTGCGCTCGGACGCAATTAAGCAATTTAACTTGACAA 2720
 859 LeuThrPlyValThrGlyAsnSerglyLeuYsLeuAsnSerThrGcl 875
 2721 CGGCACATTTACATCAATTCGCGCTATCGACACATCGGACGCGCGC 2770
 875 ySerMetValLeuPheAsn..... 881
 2771 AAACGCGAGTGGCGAGATGGCGCGCGCGCTTCGCGCGCTTCCTA 2820
 881 881
 2821 TTATCGTTAGCGCGCAACTTCGCGAGATCCGTTTCAACACGCTGAC 2870
 882GlyGlyLysAsnIlePheAsnThrLeuThr 891
 2871 GGTAAAGCGCAATTCGACGCTCAGGAAACATTCGCTTATGTCGGAAC 2920
 891 rValAspGluLeuThrThrSerAsnSerAlaPheValMetArgThrAsnT 908
 2921 TCTTGGCTACCGCAGCGCAATTCGCGAAGTTCGGAAGTTCGGAAGC 2970
 908 hr.....GlnGlnAlaAspGlnLeuIleValYsAsnLysLeuGln 922
 2971 ACTTACACCTTGCT.....GTCAACATACCGGCAACGAAC 3008
 923 AlaAsnAsnLeuLeuValAspPheIleGlnLysGlyAsnAspLys 939
 3009 CGTAGCTCTCGCAATTCGAGCTGTGAGAAAGAAACAAACACACACCGC 3058
 939 sAsnGlyLeu...AsnIleAspLeuValLysAlaProGlnAsnThr... 953
 3059 TGTCCGAATAATCTTAATTTACACCTGCAAAAGAAACACGATGCCGCGC 3108
 954 ...SerLysAspVal...PheLysThrGlnThrGlnThrIle..... 965
 3109 GCATGGCTTATCAGCTTATCGCAAGACGCGAGTTCCGCTGCATATA 3158
 966GlyPheSerAspValThrPr 972

3159 TCCGGTCAAGAAACAGAGCTTTCGCAAACTCGGCAAGCGGAGAAA 3208
 972 ogIuIleYsGlnGlnGlu.....LysAspIlyLys 983
 3209 CAGAGCGCGCTTGCAGCGCAAAAGACACAACTTCCGCGCAACACAG 3258
 983 erValTrpThrLeuThr.....GlyTyrLysThrVal 993
 3259 GCGGAAAAAGACACCGCGCAAGCTTACCGCTGATTCGCGCGCG... 3306
 994 AlaAsnAlaAspAlaAlaLysLysAlaThrSerLeuMetSerglyGlyTyr 1010
 3307CGCAATGCCACCGCAAAAGCAGAAAGTG- 3334
 1010 rLysAlaPheLeuAlaGluValAsnAsnLeuAsnLysArgMetLysSpl 1027
 3335 TTGCGGAACGCGCGCGGACGAGCG... 3360
 1027 euArgAspIleAsnGlyGluAlaGlyAlaTrpAlaArgIleMetSergly 1043
 3361 ...GGGCAAAATGCCGCTATATGACGCGGAGCAAGCAAAACGGGT 3407
 1044 ThrGlySerAlaGlyGlyGlyPheSerAspAsnTyrThrIleValGlnVa 1060
 3408 GCAGCGGATTAAGACACCGCGCTTG...G 3433
 1060 lGlyAlaAspAsnLysHisGluLeuAspGlyLeuAspLeuPheThrGlyV 1077
 3434 GGAACACGCGGAAAGCAACCGCGCGCTACACCGCTTCCCGCGC 3483
 1077 alThrMetThrTyrThrAspSerHisAlaGlySerAspAlaPheSergly 1093
 3484 GCCCGCGCGCGCGCGGATTCGCGCAACCGGACCGCAACCGCAAC 3533
 1094 GluThrLysSerValGlyAlaGlyLeuThrAlaSerAlaMetPheGluSe 1110
 3534 CCAACGCGAGCGGCGCTGATCACCGCTTATGCAATGACGCTTGTGAGTG 3583
 1110 rGlyAlaTyrIleAspLeuIleGlyLysTyrValIleHisAsp...AsnG 1126
 3584 AATTTCCGCGCAGCGCTCACAGGCTTTCGCGCTACGACGCAATTGAC 3633
 1126 lutyThrAlaThrPheAlaGlyLeu..... 1134
 3634 CGCGTGTTCGCAAGACCGCGCAACCGCTTTCGCAACGCGCATCCG 3683
 1135GlyThrAla 1137
 3684 GGACACCAAAACACTACCGTTGCAAGATTCGCGCGC... 3720
 1137 gAsp.....TyrSerSerHisSertTrpAlaGlyAlaGluValG 1151
 3721 ...TACGCGCAACAA...ACCGACCTGGCGCAATTCGCTATGCAAGAAC 3765
 1151 lYtyrAlaGlyrHisValThrAspSerAlaTrpIleGluProGlnAlaGlu 1167
 3766 CTCGCGACGCGCGCGCTC...GGCATCTGTTTTCGCAACGCGGACCG 3812
 1168 leuValTyrGlyAlaValSerGlyLysGlnPheSerTrpLysAspGlnG 1184
 3813 A.....AACACTTTCGACGAGCGCATCGCACT 3841
 1184 yMetAsnLeuThrMetLysAspLysAspPheAsnProLeuIleGlyArgT 1201
 3842 CGGACGCGCTTGGCCACGCGCTTTCGCGCAATTCGCGACG 3891
 1201 hr.....GlyValAspValGlyLysSerPheSerglyLys 1212
 3892 ...TTGACATTCGCGCATCAGCGCGCGGCTTATGAGCGGCGCT 3938
 1213 AspTrpLysValThrAlaArgAlaGlyLeuGlyTyrGlnPheAspLeuPh 1229
 3939 TTCAGACGCGC.....ATCA 3952

```

1229 ealaasnglygluthtValleuargaspalaserglyglulysargllel 1246
      ::::::::::
3953 GAGGC...AAATCCGCCGCCGCTGCTGATTCAGGCATTACGGCAACA 3999
      ::::: ||| ::::::::::: ||:::
1246 ysglygluysaspgrlyargmetleumetasValglyleuasnalaclu 1262
      :::::::::::
4000 TACCGC.....GCAAGGTTGCG 4016
      |||
1263 lIeargAspAsnValarPheglyLeuGlupheglulysSerAlahecl 1279
      :::::::::::
4017 CGGATTCGGCATCGAACCGACATCGCGCAACGCCCTATTTCGCCAAA 4066
      | ::::::::::: |||
1279 yLysTyAsnValAspAsnAlaIle.....A 1288
      :::::::::::
4067 AACGGATTACCGATAC 4083
      :::::::::::
1288 snAlaAsnPheArgTyr 1293
      :::::::::::

```

seq_name: sp_bacteria:Q48206

seq_documentation_block:

```

ID Q48206 PRELIMINARY; PRT; 492 AA.
AC Q48206;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DA66;
RC MEDLINE=95302961; PubMed=7783620;
RA Lomholt H.; Poulsen K.; Mogens K.;
RT "Comparative characterisation of the iga gene encoding IGA1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae."
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82467; CAA57850.1;
DR MEMOPS; S06.001;
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1.1.
DR PRINTS; PR00921; IGASERPPASE.
KW Protease.
FT NON_TER 1 1
FT NON_TER 492 492
SQ SEQUENCE 492 AA; 54751 MW; 6FP93038E7AD5984 CRC64;

```

alignment_scores:

```

Quality: 650.50 Length: 534
Ratio: 2.147 Gaps: 18
Percent Similarity: 56.742 Percent Identity: 32.210

```

alignment_block:

US-09-303-518D-653 x Q48206 ..

Align seg 1/1 to: Q48206 from: 1 to: 492

```

202 GCGAAGATATTGAGTTTACACAAAAGGGAGTTGGTGGCAATC 251
    ||| ::::::::::: ||:::
1 AlathrasnValGluValArgAspLysAsnAsnLeuGlySerVa 17
252 GATACGAAA...GCCCGATGATTGATTCTTCTGTGTATCG...CGTA 295
    :::: ||| ||||| ||||| ||||| ||||| ::::
17 lIeuprolyAspIleProMetIleAspPheSerValValAspValAsp 34
    :::: ||| ||||| ||||| ||||| ||||| ::::
296 ACGCGTCGCGCATTTGGCGCGCATCAATATTATTGAGCGTGACAT 345
    :::: ||| ||||| ||||| ||||| ||||| ::::

```

```

34 ySarglleAlaThrLeuIleAsnProGlnTyrValValGlyValLysHis 50
346 AACGC...GGCTATAACATGTTGATTGTG..... 375
    ||| ||| ::::::::::: |||||
51 ValGlyAsnGlyValGlyGluLeuHisPheGlyLysLeuAsnGlyAsnTr 67
    :::: ||| ::::::::::: |||||
376 ...GCGAGGGAAGCAATCCGATCCAGCACCGC..... 405
    :::: ||| |||||
67 PAsnProLysPheGlyAsnSerIleGlnHisArgAspValSerTrpGlu 84
    :::: ||| ::::::::::: |||||
406 ..TTTCTTACCAATTGTGAAGAATAATATTAAAGCAGGAGCTAAC 453
    ||| ||| ::::::::::: |||||
84 lAsnArgTyrTyrThrValGluLysAsnAsnPheSerSerGluLeuAsn 100
    :::: ||| ::::::::::: |||||
454 GCGCATCCT..... 462
    |||
101 GlyLysThrGlnAsnAsnGluLysAspLysGlnTyrThrSerAsnLys 117
    :::: |||
463 .....TATGC..... 468
    |||||
117 sAspValProSerGluLeuTyrGlyGlnAlaLeuValLysGluGlnGln 134
    :::: |||
469 .....GCGATTATCATATGCCCGCTTTCACAAATTGTGC 504
    ||||| ||||| ||||| ||||| |||||
134 snGlnLysArgArgGluAspTyrTyrMetProArgLeuAspLysheVal 150
    :::: |||
505 ACAGATGCGAAGCTGTGAG...ATGACAGATTATGATGAGTGGAA 551
    ||||| ||||| ||||| |||||
151 ThrGluValAlaProIleGluAlaSerThrThrSerSerAspAlaGlyTh 167
    :::: |||
552 ATACGCTGATTTAAATAATACCTGATCGTGTTCGAATCGAGACGCA 601
    ||| ||| ||||| ||||| |||||
167 rTyAsnAspGlnAsnLysTyrProAlaPheValArgLeuGlySerLys 184
    :::: |||
602 GACATAT.....TGCGGCTCGATGACGACGA 630
    |||||
184 erGlnPheIleTyrLysLysGlySerHisTyrGluLeuIleLeuGlu 200
    :::: |||
631 CCCAATAACCGGCAAGTTCATATCAT..... 657
    |||||
201 LysAsnGluLysGluIleIleHisArgTyrAspValGlyLysAsp 217
    :::: |||
658 .....ATTGCAAGCGCATATTCTTGCTGCTGGTGGCAATVACCT 697
    :::: |||
217 nLeuLysLeuValGlyAsnAlaTyrThrTyrGlyLysLeuAlaGlyThrProt 234
    :::: |||
698 TTGCACAAATGATCAGTGTGTCGACAGTCACTTACGTAGCCAAA 747
    :::: |||
234 yLysValAsnHisThrAspAspGlyLeuIleGlyPheLysPheThr 250
    :::: |||
748 ATTAACATGACCA..... 762
    |||||
251 GluAspHisAsnAspProLysGluIleLeuSerArgLysProLeuThrAs 267
    :::: |||
763 .TATGTTTTTTACCAACGAGGCTCATTTGCGCACAGTGGTACCAA 811
    ||||| ||||| ||||| |||||
267 nTyrAlaValLeu.....GlyAspSerGlySerProL 278
    :::: |||
812 TGTATTATGATGATGCCCAAAAGCAAAAGTGTATTATATGGGTATTG 861
    :::: |||
278 eUpheValTyrAspLysSerLysGluLysTrpLeuPheLeuGlyLys 294
    :::: |||
862 CAACACGACACCCCTATATAGAAAGCAATGGCTTCACCTAGTTCG 911
    :::: |||
295 AspPheTrpGlyGlyTyrLysLysLysSerTrpGlnGluTrpAsnIle 311
    :::: |||
912 TAAAGATTGTTCTATGATGAATCTTTCGTGAGATACCATTCAGTAT 961
    ||| ||| ::::::::::: |||||
311 rLysProGlnPheAlaGluAsnIleLeuLysLysAspSerAlaGlyLeu 328
    :::: |||
962 TCTACGACCACTCAAAATGGCAATACTTTTATAGACAAATATAT 1011
    :::: |||
328 eu.....LysGlyAsnThrGlnTyrAsnTrpThrSerLysGlyAsn 341
    :::: |||

```

1012 GCGCAGAGAAAAATCGATGCCAAACATAACTATCTCTACCTATAG 1061
 342 ThrSer..... 343
 1062 ATTAAAAACGACGACCGTTCATGTTTAAATGTTCTTATCCGAGACAG 1111
 344 .LeuIleSerGlyThrSerGlySerLeuSerValAspLeuValAsp.... 358
 1112 CAGAGAACCTGTTTATCATGCTGCGAGGTGGGTCAACAGTTTCACACC 1161
 359AsnLys 360
 1162 AGACTGATATATGAGAAAAATATTTCTTATGACAAAGAAAGTGA 1211
 361 AsnLeuAsnHisGlyLysAsnValThrPhe.....GluGlySerLys 375
 1212 ATTGATCTTACGACGACATCAACAGAGCGCGCGGTGTGTTATTTG 1261
 375 .IleuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyLysLeuPheG 392
 1262 AGGTAATTTTACGTC...TCGCTAAAAACAAGAACGAGGAGGAGC 1308
 392 .IleLysAspLysGlyValLysGlyThrSerGluAsnThrThrTrpLysGly 408
 1309 GCGGCGCTTCATATCATGATGACGACGACGCTTACTTGAAGTAAACG 1358
 409 .AlaGlyIleSerValAlaGlnGlyLysThrValLysTrpLysValHis 425
 1359 CGTGGCAACGACCGCTGTCGCAAAATCGCAAGGACGCGTGTGTC 1408
 425 .nProGlnPheAspArgLeuAlaLysIleGlyLysGlyLysLeuIleVal 442
 1409 AAGCCAAAGGGAACCAAGCGCTGCGTACGCGTGGGAGCGTAAAGTC 1458
 442 .IleLysArgLysAspAsnLysGlySerLeuLysValGlyAspLysTrpVal 458
 1459 ATCTTGATCAGCAGCGGACGATCAAGCAAAACAAGCCTTAAAGTA 1508
 459 .ValLeuLysGlnGlnThrThr.....GlyGlnHisAlaPheAla 473
 1509 AATCGCTTGTCAGCGGAGGCGGAGCGTGCACATGATGCCGATATC 1558
 473 .ValGlyIleValSerGlyArgSerThrValIleuAsnAspAspAsnG 490
 1559 AG 1560
 490 In 490
 seq_name: sp_bacteria:Q9F6T1
 seq_documentation_block:
 ID Q9F6T1 PRELIMINARY; PRT; 1295 AA.
 AC Q9F6T1;
 DT 01-MAR-2001 (Tremblere, 16, Created)
 DT 01-MAR-2001 (Tremblere, 16, Last sequence update)
 DT 01-DEC-2001 (Tremblere, 19, Last annotation update)
 DE SECRETED AUTOTRANSPORTER TOXIN.
 GN SAT.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CF703;
 RX MEDLINE=20485608; PubMed=11029690;
 RA Guyer D.M., Henderson I.R., Nataro J.P., Mobley H.L.;
 RT "Identification of Sat, an autotransporter toxin produced by
 RT uropathogenic Escherichia coli";
 RL Mol. Microbiol. 38:53-66(2000).
 DR EMBL: AF289092; AAC30168.1; -
 MEROPS: S06.004; -

DR InterPro: IPR000710: IGA_S6.
 DR Pfam: PF02395; IGA1; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 SO SEQUENCE 1295 AA; 140042 MW; 81122069273B8F5 CRC64;
 Alignment_scores:
 Quality: 635.00 Length: 1470
 Ratio: 0.893 Gaps: 68
 Percent Similarity: 48.367 Percent Identity: 23.605
 alignment_block:
 us-09-303-518d-653 x Q9F6T1 ..
 Align seg 1/1 to: Q9F6T1 from: 1 to: 1295
 160 TATCGCAGCTTTCGCGAAATTAAGCAAGTTTCAGTCGCGCGCAAGA 209
 62 TyrLeuAspLeuAlaGlnAsnLysGlyIlePheGlnProGlyAlaThrAs 78
 210 TATTCAGTTTACCAACAAAAGGAGAGTTGTCGCGCAATCGTAGCA 259
 78 .PvalThrIleThrLeuLysAsnGlyAspLysPheSerPheHisAsnLeu 95
 260 AAGCCCGATGATGATTTTCTGCTGATGCGCTAAGCGGCGGCGCA 309
 95 .erLlePro.....AspSerSerGlyAlaAlaAlaSerGlyAlaAlaThr 109
 310 TTGCGGCGGATCAATATATGAGCGTGGCAGATACGCGGCTTAA 359
 110 .AlaIleGlyGlySerLysSerValThrValAlaHisAsn..... 122
 360 CAATGTGATTTTGGTGGGAGGAGCAATCCCGAT..... 396
 123LysLysAsnProGlnAlaAlaGlnThrG 132
 397CAGCAGCGCTTTCCTTACCAAAATGTAAGAAATTAATATAA 441
 132 .InValTyrAlaGlnSerSerLysArgValAlaAspArgAsn..... 146
 442 GCAGGAGCTAACGGCCATCTTATGCGCGGATTAATATGCCGCTT 491
 147SerAsnAspPheGlnIleGlnArgLe 155
 492 GCACAAATTTGTCACAGAT.....GCAGAACCTGTGAGATG.... 528
 155 .uAsnLysPheValAlaGlnThrValGlyAlaThrProAlaGlnThrAsn 172
 529 .ACCACTTATATGAT.....GGTGGAAATACGCTGAT 561
 172 .roThrThrLysSerAspAlaLeuGlnArgLysGlyIleValThrSerAsp 188
 562 TTAATTAATTAACCTGATCGTTCGATCGAGCAGCGCAGACAAATTTG 611
 189 .GlySerLysLysIleIleGlyPheThrGlyLysSerGlyLysThrSerP 205
 612 CGCGTGTATGAACAGCAACCAATATACCGCAAGTTCATATCATATG 661
 205 .e.....IleAsnGlyLysSerLysIleSerThr 215
 662 CAGCGCATATTTCTTG...CTGCTGCTGGCATATACCTTTCACAAAT 708
 215 .snSerAlaTyrSerHisAspLeuLeuSerAlaSerLeuPhe..... 228
 709 GGATCAGGTGTGGCAGCTCACTTAGTAGCGAAATTAACATAG 758
 229GluValThrGlnThrPAs 234
 759 CCAATATGCTTTTACCAACAGAGAGCTCA..... 789
 234 .pSerTyrGlyMetMetIleTyrLysAsnAspLysThrPheArgAsnLeuG 251
 790TTTGGCAGACGTGCTCACCAATGTTTATCATGATGCCCAAAAG 834

```
|||||.....:
251 luIlePheGlyaspSerGlyalaTyLeuTyraSpasnyLseu 267
835 CAAGAAGTGTAAATTAAGGGGTATTGCAACAGCAGCAACCCCTATATGG 884
:::|||||.....:
268 GluTySPValleuValGlyThrHisGly.....LeuI 280
885 AAAAAAGCATGGCTTCAGCTAGTCTGTAAGATTGTTCTATGATGAAA 934
:::|||||.....:
280 aserValaspGlnleuThrTrpIleThrIysTyraSpasnySL 297
935 TCTTTGCT.....GAGATACCATTCAGTATTCACAAACCACATCAA 978
:::|||||.....:
297 euValSerGlyleuLysaspThrTySer..... 306
979 AATGGGAATACTTTTAAAGCAATATATATGGCGAGAAAAATGGA 1028
|||.....:
307 ..HisLysIleasnleuasnGlyasnSn..... 315
1029 TGCCAAACATAACACTATTCTCTACCTTATGATTAAAAACAGCAACG 1078
316 .....ValThrIleLysasnTrpAspi 323
1079 TTCAATGTGTTAATGTTCTTTATCCGAGACAGCAAGAGAACCTGTTAT 1128
:::|||||.....:
323 leThrleuHisGlnasnSnalaAspThrGlyThrGln..... 336
1129 CATGCTGCAGGTGGGTCAACAGTTATGACCCAGACTGAATATGAGA 1178
337 .....GluLysIleThrLysaspLy 343
1179 AATATTTCTTTATTATGACAAAGAAAGTGAATTGATTACTTACCAGCA 1228
:::|||||.....:
343 sasplleValalPheThrasn.....GlyGlyaspValleuPheLysaSpa 358
1229 ACATCAACCAAGCGCGGGGGTGTGATTTT...GAGGGAATTTTACG 1275
|||.....:
358 snleuaspPheGlySerGlyIleIlePheaspGlyGlyHisGlyLys 374
1276 GTCTGGCCTAACAAACAGCAACCTGGCAAGCGCGGGCTTCATATGAG 1325
:::|||||.....:
375 AsnIleasnGlyGlnGlyPheThrPheLysGlyalaGlyIleaspIleG 391
1326 TGATGGACGTACCTTACTTGAAGTAACGGCGTGGCAAAAGCAGCCGC 1375
|||.....:
391 yLysGlySerIleValasnTrpAsnalaLeuTyraSerSeraspValL 408
1376 TGTCCAAAATTCGGCAAAAGCAGCTGTGTTCAAGCCAAAGGGGAAAAC 1425
|||.....:
408 euHisLysIleGlyProGlyThrleuAsnValGlnLys..... 421
1426 CAAGGC...TCGGTCAAGCTGGCGCAGCGTAAGCATCTTAGATCACA 1472
|||.....:
422 GlnGlyalaAsnIleLysIleGlyGlnLysnValIleLeuasnGlnG 438
1473 GCGGACGATCAAGCAAAAAACAAGCCTTAGTAATCGCGTGTGCA 1522
:::|||||.....:
438 uGly.....ThrPheasnSnIleTyraLeuAlas 448
1523 GCGGACAGGGGAGGTGCAGTAATGCCGATATCACTCAACCCGAG 1572
|||.....:
448 ergLysasnGlyValIleLeuasnLysaspasSerleuGlnLysnsp 464
1573 AAA.....CTCTATTGGCTTGGCGGCGGAGCTTTGGATTGAA 1613
:::|||||.....:
465 GlnTyraIleGlyIlePhePheThrLysaspGlyGlyThrleuaspLeu 481
1614 CGGCGATTCGCTTGTCCACCGCATTCAAAATACCGATGAAGGGGGA 1663
|||.....:
481 nGlyHisasnGlnThrPheThrArgIlealaIaThrspaspGlyThrT 498
1664 TGAATGTCAACACATCAAGCAAGAATCCACCGTTACCATTCAGAGC 1713
|||.....:
498 hrIleThrasnSeraspThrThrLysGlnalaValleuAlaIle..... 512
1714 AATAAAGATATTACTACACCGGAATACACAACTGGATGACAAAAA 1763
513 .....AsnSnleuLysSer..... 517
1764 AGAAATTGCCTACAAACGTTTGGTTGGCGAGAAAAGATGACAAACGA 1813
|||.....:
518 ....TyrlleThrHisGly.....AsnIleA 525
1814 ACGGCGCGCTC.....AATCTGAATTACACACCGGAAGAGCG 1851
|||.....:
525 snGlyasnIleLysleuThrHisasnIleasnSerGlnaspLysThr 541
1852 GATCGACCTTACGCTTCCGGCGGCAAAATTTAAAGCGCAATATCAC 1901
:::|||||.....:
542 AsnalaLysleuIleLeuaspGlySerValasnThrLysasnaspValG 558
1902 GCAAAACAAGCGCAACCTGTTTTCAGCGGCGAGACCGCAGCGCT 1951
|||.....:
558 uValSerasnalaSerleuThrMetGlnGlyHisAlaThrGlnHisAla 575
1952 AC.....AATCAT.....TTAGAAGC 1968
575 lePheArgSerSerAlaasnHisCysSerleuValPheLeuGlyThr 591
1969 GCGTGG..... 1974
592 AspTrpAlaThrValleuLysGlnThrGlySerSerTyraSnLysSpH 608
1975 .....TCAAAAATGGAAGTATCCACACAGAGAAATGTTGGGACACAG 2020
608 easnSerAspTrpLysSerasnGlnGlnThrSerPheaspGlnPro 625
2021 ATTGATTCGACCGGCATTTAAACGGGAAACTTCATTCAGGCGGA 2070
|||.....:
625 sprTrpLysThrGlyValPheLysPheaspThrleuHisleuAsnAla 641
2071 CAACGGGTGTTTCCGCAATGTTGCCAAGTGGAAAGCCATTCGCAATT 2120
:::|||||.....:
642 AspPheSerIleSerArgasn...AlaasnValGlnLys.....asnI 655
2121 AAGCAATCAAGCCCAAGCACTTTCGGTTCGCCACGCATCAAAAGCCACA 2170
|||.....:
655 eserAlaasnLysSerAlaIle..... 662
2171 CAATCTGTACAGTTCGGACTGACGGGTCTGACAACTGTACCGAAAAA 2220
662 ..... 662
2221 ACCATTACCGACGATAAAGTG...ATTGCTTCATTGAGCAACCGCAT 2267
|||.....:
663 ThrIleGlyaspLysasnValTyrlleaspasnleuAlaGlyLysasnI 679
2268 CAGAGCAAAATGTACGCTTCCGATCAGCTCATTTAAATCTCACAGAG 2317
|||.....:
679 eThrAsnAsn.....GlyPheaspPheL 687
2318 TTGCCACACTCAACGCAATCTTAGTCAGCGGAGACACGCACTATACG 2367
|||.....:
687 yscGlnThrIleSerThrAsnleuSerIle...GlyGlnThrLysPheThr 702
2368 .....GTTACGGCGCAAGCGCCAAAGCGCAACCTCAGCCGCGGG 2411
703 GlyGlyIleThr.....AlaHisasnSerGlnIleAlaIleGlyas 716
2412 CAATGCCCAAGCAACATTTAATCAAGCGCACATTTAAAGCGCAACATCG 2461
|||.....:
716 pGlnalaValValThrleuasnGlyAlaThrPheleuaspasThrPro 733
2462 CTTCG...GACAAATGCTTCATTTAATCTAAGCAACAAGCCGTA..... 2502
|||.....:
733 leSerIleaspLysGlyAlaLysValIlealaGlnasnSerMetPheThr 749
```



```

1282 AsnValAspAsnAlaLe.....AsnAlAs 1290
      :::::  |||  :::::
4074 TTACCGATAC 4083
      :::::|||||
1290 npheargTyr 1293

```

385 LeuaspIleGlyAspAsnThrThrValAspTrpAsnValLysGlyValIva 401
1365 AAGACGCGCGCTGTCCAAATCGCAAGGACGCTGTGTTCAGGCA 1414
401 IGIYAspAsnLeuHisLysIleGlyAlaGlyThrLeuAsnValAsnValS 418
1415 AAGGGAAACCCAAAGGCTGCGTCAAGCGGACGGAATGATCTTA 1464
418 erGlnGlyAsn.....AsnLeuSerThrGlyAspGlyLeuValValLeu 432
1465 GATCAGCAGCGCGAGATCAAGGCAAAACAGCCTTACTGAATCGG 1514
433 Asn.....SerAlaAsnAlaPheAspAsnIleTyr 442
1515 CTGTGTCAGCGGACGAGGCGGCTGCACCTGAT.....GCCGATA 1555
442 rMetLAspSerGlyHisGlyValValLysIleAsnHisSerAlaAlaLeuA 459
1556 ATCAGTTCAACCCCGACAAA...CTCTATTTCGGCTTCGGCGGACGT 1602
459 snGlnAsnAsnAspTyrAlrGlyIlePhePheThrGluAsnGlyLThr 475
1603 TTGGATTGACGGGCGCATTCGCTTCCTTCACCGCATTCAAATACGA 1652
476 LeuAspLeuAsnGlyTyrAspGlnSerPheAsnLysIleAlaAlaThrAs 492
1653 TGAAGGGCGGATGATGTGCACCAATCAAGACAAAGATCCACCGCTA 1702
492 pIleGlyAlaLeuIleThrAsn.....SerAlaVal. 502
1703 CCATTACAGGCAATTAAGATATTACTACACCGGCAATACACACATTTG 1752
503GlnLysAlaValLeuSerValAsnAsnGlnSerAsn... 514
1753 GATACAAAAAAGAAATTCCTACACGCTTGTTGGCGAGAAAGATGC 1802
515TyrMetLysIleGly.....Se 520
1803 AACCAAAAGACGAGGCGGCTCATCTGATTCACACCGGAAAGAGCG 1852
520 rValSerGlyAsnThrGluIleAsnHisGlnPheAspThrGlnLysAsnA 537
1853 ATCGCACTTACTGCTTCGCGGAGACAAATTTAAAGCGCAATTCACG 1902
537 snSerAlrGleuIleLeuAspGlyAsnValAspIleThrAsnAspIleAsn 553
1903 CAAACAAAGCGCAACTGTTTTCAGCGGACGACACCGCACCGCCTA 1952
554 IleLysAsnSerGlnLeuThrMetGlnGlyHisAlaThrSerHisAlaVa 570
1953 CAATCATTTAGGAGCGGCTGTCAAAAATGGAAGTATC..... 1992
570 lPheAlrGlnGlyValThrCysMetLeuProGlyValIleCysGluL 587
1992 1992
587 yAspTyrValSerGlyIleGlnGlnGlnGluAsnSerAlaAsnLysAsn 603
1993CCACAGAGCAAAATCGTGTGGACAA 2018
604 AsnAsnThrAspTyrLysThrAsnAsnGlnValSerSerPheGlnGlnPr 620
2019 CGATTGATCGACCGCACATTTAAAGCGGAAATTCATATTTCAGGGCG 2068
620 oAspTrpGluAsnAlrGleuPheLysPheLysThrLeuAsnLeuIleAsnS 637
2069 GACAAAGCGGTGTTCCGCAATGTTCGCAAAAGGAGCGATTCGCAT 2118
637 erAspPheIleValGlyAlrAsn...AlaIleValValGlyAspLieser 652
2119 TTAAGCAATCAGCCCAAGCAGTTTTCGTGTGCGACCGCATCAAGCA 2168
653 AlaAsnAsnSerThrLeuSerLeuSerGly..... 662

2169 CACAACTGTACACGTTTCGAGCTGACGCGGCTGACAACTGTTCAGCAA 2218
663LysAspThrLysValHisIleAspMetTyrAspGlyL 675
2219 AAACATTTACC.....GACGTTAA 2238
675 yAsnIleThrGlyAspGlyPheGlyPheAlrGlnAspIleLysAspGly 691
2239 GTGATTCCTTCATTGACCAAGCGACATCAGGCGCATGTGACCTTCG 2288
692 ValSerValSerProGlnSerSerTyrPheGlyAsnValThrLeuAs 708
2289 CGATCAGCGCTTAATTCCTACAGACTTGCACACTGCAACGCAATC 2338
708 nAsnHisSerLeuLeuAspIle.....GlyAsnL 718
2339 TTAGTCAGCGGCA.....GACAGCAGCTATACGTTACGCGC 2376
718 yAspThrGlyGlyIleGluAlaTyrAspSerSerValSerValThr... 733
2377 AACGCCACCAAAAGCGCAACTCAGCTCGTGTGCGCAATGCCAAGCAAC 2426
734SerGlnAsnAlaValPheAspAlrGly..... 743
2427 ATTTAATCAAGCCACATTAACGGAACACATCGGCTTCGACATGCTT 2476
744SerPheValAsnSerSerLeuThrLeuGluLysGlyAlaL 757
2477 CATTTAATCTAACCAACACGCGGTACAAAAGCGAGCTGCAGCTTTC 2526
757 ySerLeuThrAlrGlnGlyLysIlePheSerThrGlyAlaValAspValLys 773
2527 GACACGCTAAGGCAAAAGCTAAGCCATTCGCACTGACAGCGCATGTCTC 2576
774 GluAsnLAspLeu.....IleLeuThrGlyThrProSe 785
2577 CTTAGCCGATTAAGGCAATTCATTTGAAACAGCCGCTTACCGGAA 2626
785 r...AlaGlnLysGlnGluLyrThrSerProValIleSerThrThrGluG 801
2627 AAATCAGCGCGCGCAAGATACGCGCATTCACCTTAAAGACAGCGAATG 2676
801 LysIleAsnLeuGlyAspLysAlaSerLeuSerValLysAsnMetGlyTyr 817
2677ACGTCGCGTGGCGACGCAATTAAGCAATTAACCTTA 2717
818 LeuSerSerAlrIleHisAlaGlyThrThrAlaAlaThrIleAsnLeuG 834
2718 CAACGCCACCATTAACACTCAATTCGCTTACGACAGATGCGGAGCG 2767
834 yAsp.....GlyAspA 838
2768 CGCAACCGCGCAGTGGCGGAGATCCGCGCGCGCTTCGCGCGCTTC 2817
838 lAlrLThrAspSerProLeuPheSerLeuMetLysGlyTyrAsnAla 854
2818 CTATTATCC.....GTTACGCGCGCAACTTCGCGCAATCCGTTTCA 2861
855 ValLeuSerGlyAsnIleThrGlyGlnGlnSerThrValAsnMetAsnS 871
2862 CAGCGTCAGCGTAAAGCGCAAAATGGAAGGTCAGGACACTTCGCTTA 2911
871 nAlaLeuTrpTyrSerAspLysAsnSerThrIleGlyThrLeuLysSerT 888
2912 TGTG.....GACCTCTGCGC.....TACGCGACGCGCA 2943
888 hrGlyGlyAlrGlyValGlnLeuGlyLysAspPheAlrThrLeuAlr 904
2944 TTGAAGCTGGCGCAAGTTCGGAAGGCACTTACACTTGGCTGTCAACA 2993
905 ValLysGluLeuAsnAlaAsnAlaThrPheLeuMetHisThrAsnS 921


```

2994 TACGGGACGACCGCGTAGTCTCGAGCAATTGACGTTAGTGAA.... 3039
    |||:
921 nser.....GlnAlaSpGlnLeuSsnValThrAsnLysL 933
3040 .....GGAAAGACACACACCGCGTGTCCGAAATCTTAATTTCACCCG 3084
    |||:||||| ||| ||| ||| ||| |||
933 euleuGlySerAsnSsnThrValLeuValAspPheLeuSsn..... 946
3085 CAAGACGACGCTGCATGCCGCGCATGCGCTTATCAGCTTATCCGGA 3134
946 ..... 946
3135 AGACGGCGAGTCCGCTGCATATCCGCTCAAGACAGAGCTTTCCG 3184
947 .....LysProAlaSerGlnMetAsnValThrL 956
3185 ACAAACTCGGACGAGGAGGAGAAACAGAGCGCGCTTACGCGCAAAACG 3234
956 euIlleThrAlaProLysGlySerAspGlnLysThrPheThrAlaGlyThr 972
3235 GCACAACTTGGCGCCCAACACAGCGGAGAAAGACAGCGGCAAGGCT 3284
973 GlnGlnIleGlyPheSerAsnValThrProValIleSerThrGlnLysThr 989
3285 TGACGCG.....CTGATTGGCGCGCGCGCGCATGCCACCGAAA 3322
989 rAspAspAlaThrIleLysThrPheLeuThrGlyLysGlnThrValSerAsp 1006
3323 AGCGAGAAAGTGTGCGCAACCGCGCGCGCGCATGCCACCGCT 3372
1006 lAglyAlaSerLysThrAlaThrAspPheMetAlaSerGlyTyrLysSer 1022
3373 GGCATTATGACGCGGAGAGAAAGACAAACGCGGTGCGAGCGGATYAGA 3422
1023 PheLeuThrGlnValAsnAsnLeuAsnLysArgMetGlyAspLeuArgAs 1039
3423 CACGCGCTTGGGAGAACGCGGAGGAGAAACCGCGCGCTACCAACG 3472
1039 PThr..... 1040
3473 CTTCCCGCGCGCGCGCGCGCGCGGATTTGCGCAACCGCGAGCC 3522
1041 .....Gln.GlyAspAlaGlyVal..... 1046
3523 CAACGCAACCCCAACGCGAGCGCGCATGATCAGCGTTATGCCAATAG 3572
1046 ..... 1046
3573 CGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTCCGCGTACAG 3622
1046 ..... 1046
3623 ACAGATTGACCGCGGTGTTGCGAAGACCGCGCAACCGCGTTGACA 3672
1047 .....TrrPalaArgIleMetAsnGlyThr.....GlySe 1056
3673 AGCGGATCCGGG...ACACCAACACTACGCTTGCAGAGATTTCGCGC 3719
1056 rAlaAspGlyGlyTyrSerAspAsnLysThrHisValGlnIleGlyAla 1073
3720 CTACCGCCCAACACGAGCTGCGCAATCGTATGCAAAAACCTCG 3769
1073 sp.....ArgLysHisGln 1077
3770 GCAGGCGCGCGCTGCGCATCTGTTTGCACAAACCGGAGGAAACAC 3819
1078 LeuAspGlyValAspLeu.....PheThrGlyAlaLeuLeuTh 1090
3820 TTCGACGACGCGATCGCACTCGGACGCGCTTCCACGCTGCCGTTT 3869
1090 rTyrThrAspSerAsnAlaSerHisAla.....PheS 1102
3870 CGGCAATACGCAATCGGAGTTTCGACATCGCATCAGCGCGCGCGG 3919

```

```

|||||:||||| |||:||||| |||: |||:|||||
1102 eArgLysThrLysSerValGly..Gly..GlyLeuTyrAlaSerAlaL 1117
3920 GTTTTGATACGGCGAGCGCTTTCAGACGCGATAGAGCAAAATCCGCGC 3969
|||||:||||| ||| ||| ||| ||| |||
1117 euPheAspSerGlyAlaTyrPheAspLeuIle..GlyLys..... 1129
3970 CGCGCTGCATTCAGGCAATTCAGCAAGATACCGCGAGTTTGGCGG 4019
|||||: ||| ||| |||:||||| |||
1130 ...TyrLeuHisHisAspAsnGln.....TyrThrAlaSerPheAlaSe 1143
4020 ATTGCGCATCGAAGCGCACATCGGCGCAACGCGCTATTTCTCCAAAG 4069
|||||: ||| ||| ||| ||| ||| |||
1143 rLeuGlyThrLysAspTyrSerSerHisSerThrPyrAlaGlyAlaGly 1160
4070 CGGATTACGATACGAAACGTCATATCGCCACCGCGCGCTTGCATTC 4119
|||: ||| ||| |||
1160 aGlyTyrArgTyr..... 1164
4120 AACCGCTACCGCGCGCGCATTAAGCAGATTTATTCATTCAAACGCGCA 4169
1164 ..... 1164
4170 ACACATTTCC.....ATCAGCGCTTATTTAGCGTGTCTATA 4207
|||||: ||| ||| |||:||||| |||
1165 .HisLeuSerGlnGlySerThrPyrValGlyProGlnMetGlyLeuValLysr 1180
4208 CGGATCGCGCTTCCGCAAAAGTCCGAAACGCGCGCATACCGCGCTATTG 4257
|||||: ||| ||| ||| ||| ||| |||
1181 ..GlySerValSerGlyLys..... 1186
4258 GCGCAGATTTGCGCAAAACCGCGAGTGCAGAAATGGGCGCTAAACCGCA 4307
1187 .....SerPheSerTrp.....G 1191
4308 AATCAAGGTTTTCAGCGTGTCCCTC 4332
|||: ||| ||| ||| ||| ||| |||
1191 uAspArgGlyMetAlaLeuSerMet 1199

```

